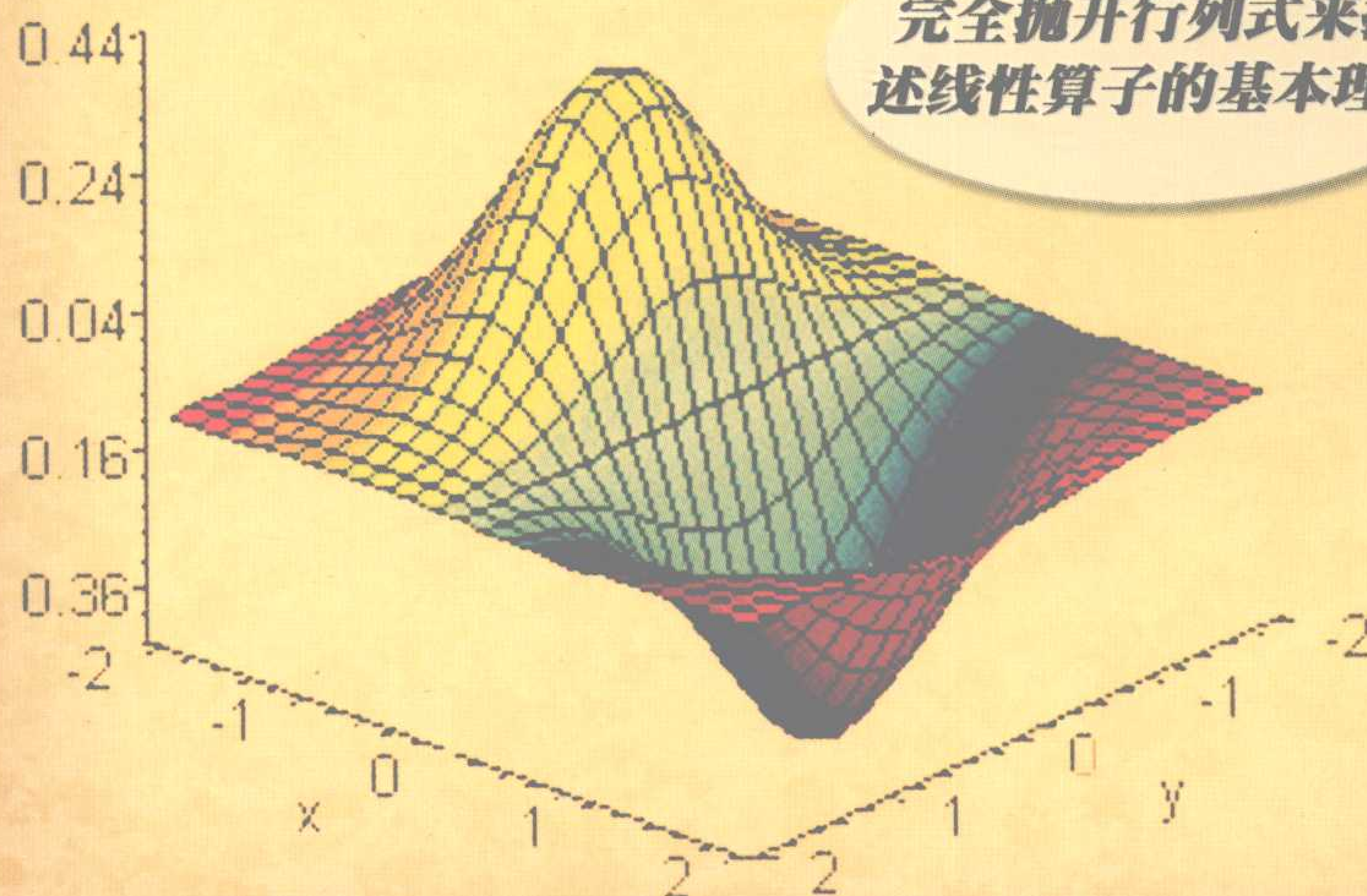
TURING 图灵数学·统计学丛书 34

Springer



完全抛开行列式来描述线性算子的基本理论

Linear Algebra Done Right

线性代数应该这样学

（第 2 版）

[美] Sheldon Axler 著

杜现昆 马晶 译



人民邮电出版社
POSTS & TELECOM PRESS

线性代数应该这样学（第2版）
Linear Algebra Done Right

“近年来最具创新性的线性代数教材，每一位大学生都不可错过。”

—— *CHOICE*

“采用完全抛开行列式的方式之后，原本曲折晦涩的证明变得优雅和直观了。”

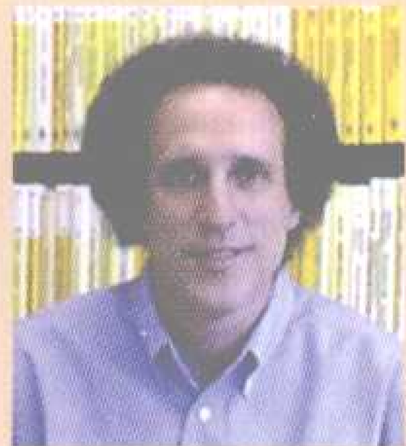
——《美国数学月刊》

“总之，本书真是一部循循善诱的杰作。”

——《数学公报》

描述线性算子的结构是线性代数的中心任务之一，传统的方法多以行列式为工具，但是行列式既难懂又不直观，其定义的引入也往往缺乏动因。本书作者独辟蹊径，抛弃了这种曲折的思路，把重点放在抽象的向量空间和线性映射上，给出的证明不使用行列式，更显得简单而直观。本书把行列式的内容放在了最后讲解，开辟了一条理解线性算子结构的新途径。书中还对一些术语、结论、证明思路、提及的数学家做了注释，增加了行文的趣味性，便于读者掌握核心概念和思想方法。

本书起点较低，不需要太多预备知识，而且特色鲜明，是公认的阐述线性代数的经典佳作。原书自出版以来，迅速风靡世界，在30多个国家为200多所高校所采用，其中包括斯坦福大学和加州大学伯克利分校等著名学府。



Sheldon Axler 1975年毕业于加州大学伯克利分校，现为旧金山州立大学理工学院院长。《美国数学月刊》编委，*Mathematical Intelligencer*主编，同时还是Springer的GTM研究生数学教材系列等多个系列丛书的主编。

延伸阅读

Lax 线性代数及其应用（第2版）

Lay 线性代数及其应用（第3版・修订版）

Golub, Van Loan 矩阵计算（第3版）

Horn, Johnson 矩阵分析（英文版），卷1和卷2

Demmel 应用数值线性代数

Trefethen, Bau 数值线性代数





本书相关信息请访问：图灵网站 http://www.turingbook.com

读者热线：(010)51095186

反馈/投稿/推荐信箱：contact@turingbook.com

分类建议 数学/线性代数

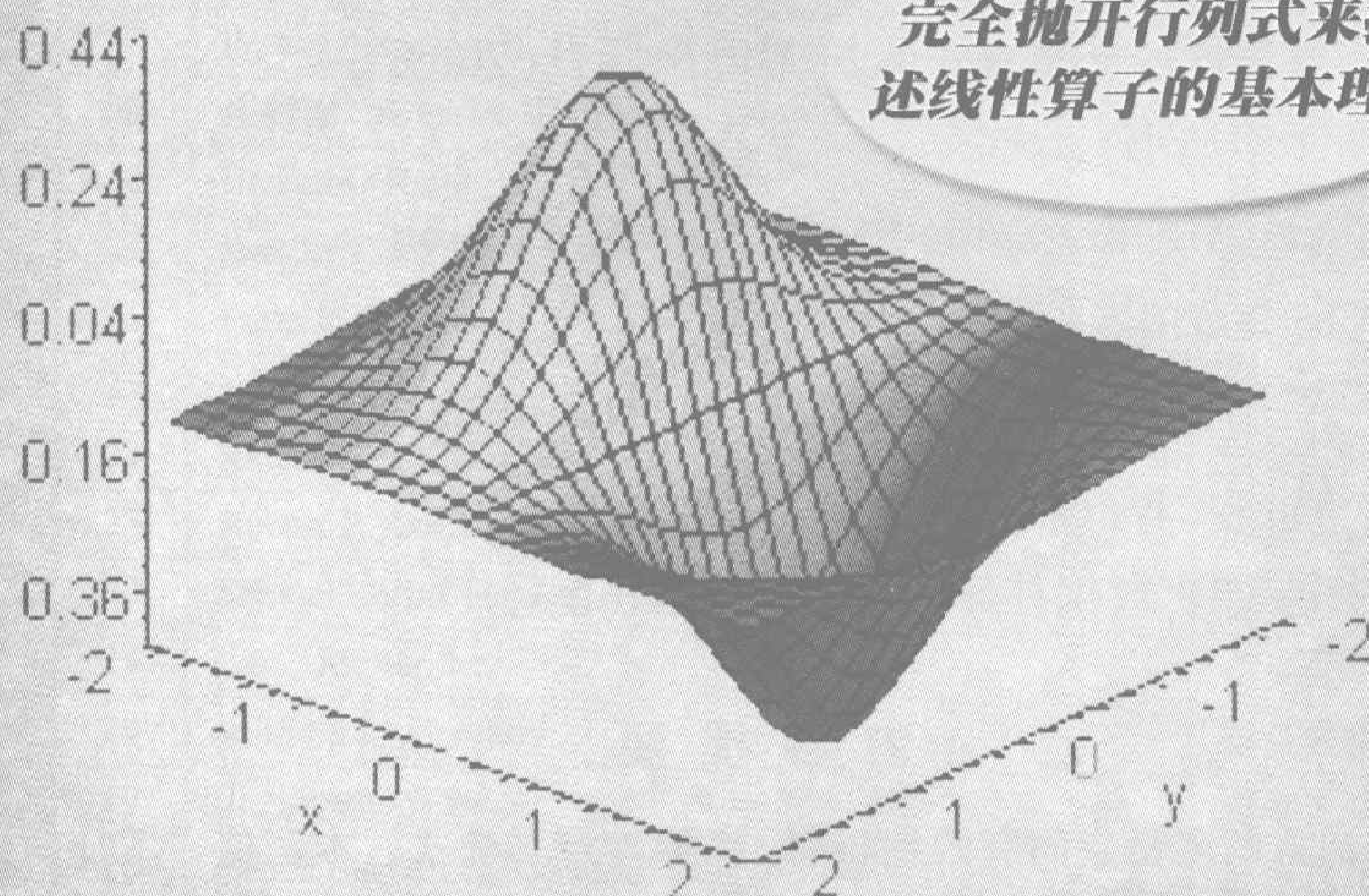
人民邮电出版社网址 www.ptpress.com.cn

ISBN 978-7-115-20614-5/O1

定价：39.00元

TURING 图灵数学·统计学丛书 34

Springer



Linear Algebra Done Right

线性代数应该这样学

（第 2 版）

[美] Sheldon Axler 著

杜现昆 马晶 译

人民邮电出版社

北 京

图书在版编目(CIP)数据

线性代数应该这样学：第 2 版 / (美) 阿克斯勒 (Axler, S.) 著；杜现昆，马晶译. —北京：人民邮电出版社, 2009. 6

(图灵数学·统计学丛书)

书名原文: Linear Algebra Done Right, Second Edition

ISBN 978-7-115-20614-5

Ⅰ. 线··· Ⅱ. ①阿···②杜···③马··· Ⅲ. 线性代数-高等学校-教学参考资料 Ⅳ. O151.2

中国版本图书馆 CIP 数据核字(2009)第 044160 号

内 容 提 要

本书强调抽象的向量空间和线性映射，内容涉及多项式、本征值、本征向量、内积空间、迹与行列式等. 本书在内容编排和处理方法上与国内通行的做法大不相同，它完全抛开行列式，采用更直接、更简捷的方法阐述了向量空间和线性算子的基本理论. 书中对一些术语、结论、数学家、证明思想和启示等做了注释，不仅增加了趣味性，还加强了读者对一些概念和思想方法的理解.

本书起点低，无需线性代数方面的预备知识即可学习，非常适合作为教材. 另外，本书方法新颖，非常值得相关教师和科研人员参考.

图灵数学•统计学丛书

线性代数应该这样学（第 2 版）

◆ 著　[美] Sheldon Axler
译　杜现昆　马　晶
责任编辑　明永玲

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街 14 号
邮编 100061　电子函件 315@ptpress.com.cn
网址：http://www.ptpress.com.cn
北京隆昌伟业印刷有限公司印刷

◆ 开本：700×1000　1/16
印张：16.5
字数：321 千字　2009 年 6 月第 1 版
印数：1–3 000 册　2009 年 6 月北京第 1 次印刷

著作权合同登记号　图字：01-2008-2562 号

ISBN 978-7-115-20614-5/O1

定价：39.00 元

读者服务热线：(010)51095186　印装质量热线：(010) 67129223

反盗版热线：(010)67171154

译者简介

杜现昆　河南省内黄县人, 生于 1961 年 8 月. 1988 年于吉林大学数学所获得博士学位. 现任吉林大学数学学院教授, 博士生导师, 吉林省数学会常务理事,《吉林大学学报理学版》编委. 主要研究环的结构理论及 Jacobi 猜测等. 著有《高等代数》(与牛凤文、原永久合著. 高等教育出版社, 2006).

马晶　辽宁省沈阳市人, 生于 1978 年 12 月. 2005 年于吉林大学数学所获得博士学位. 2005 年 9 月至 2007 年 9 月在山东大学数学与系统科学学院从事博士后研究工作. 现任吉林大学数学学院副教授, 主要从事代数学和数论方面的研究.

译 者 序

线性代数教材非常多, 这本 *Linear Algebra Done Right* 肯定是最具特色、最流行的线性代数教材之一. 原书 1996 年出版了第 1 版, 1997 年出版了第 2 版, 到 2008 年已经印刷了 11 次, 被 30 多个国家的 220 多所高等院校用作教材.

本书的主要内容是向量空间和线性算子. 描述线性算子的结构是线性代数的中心任务之一, 传统的方法多以行列式为工具. 作者认为行列式既难懂, 又不直观, 还缺少动机, 致使思路曲折, 从而掩盖了线性代数的本质. 因此, 本书完全抛开行列式, 采用更直接、更简捷的方法阐述了线性算子的基本理论, 这种独特的方法可以帮助学生更加深刻地理解线性算子的结构. 作者认为这才是恰当的方法.

本书虽然是线性代数的第二课程的教材, 但是它起点低, 由浅入深, 论述详细, 无需线性代数方面的预备知识即可学习, 因此很适合作自学教材和参考书. 本书还对一些术语、结论、数学家、证明思想和启示等做了注释, 这不仅增加了趣味性, 而且加强了读者对一些概念和思想方法的理解.

本书的内容大致相当于我国高校数学院系高等代数课程一个学期的教学内容. 本书在内容编排和处理方法上与国内通行的做法大不相同, 是一本很好的参考书, 对我国高等代数课程的教学、教研、教改都有很好的借鉴作用.

在本书的翻译过程中, 我们得到了作者 Sheldon Axler 教授及吉林大学研究生郭宏博和李建涛的帮助, 特此致谢.

由于译者的中文和英文水平都有限, 因此译文难免有词不达意之处, 欢迎读者指正.

译 者

2008 年 10 月

致　教　师

你可能正准备给学生讲授第二门线性代数课程. 学生在第一次接触线性代数时, 可能只是学习了欧氏空间和矩阵, 而现在本课程强调抽象的向量空间和线性映射.

需要对本书这个大胆的书名 (*Linear Algebra Done Right*) 做一些解释. 几乎所有的线性代数书都使用行列式来证明: 有限维复向量空间上的线性算子都有本征值. 但是, 行列式既难懂又不直观, 而且其定义的引入也往往缺少动机. 为了证明复向量空间上本征值的这个存在性定理, 大部分教科书都必须先定义行列式, 再证明一个线性映射不可逆当且仅当它的行列式等于 0, 然后定义特征多项式. 这种曲折的 (折磨人的?) 思路根本不能让学生领会为什么本征值一定存在.

与此相反, 本书所给出的不使用行列式的简单证明更直观. 本书把行列式放在了最后, 这就开辟了一条通往线性代数的主要目标 —— 理解线性算子结构的新途径.

本书从线性代数的初步知识讲起, 除了一般的数学基础之外，不再需要更多的预备知识. 本书大部分习题都需要理解书中的证明, 即使学生已经接触过本书前几章中的一些内容, 他们也会不习惯做本书所提供的这种类型的习题.

- 第 1 章给出了向量空间的定义, 并且阐述了它们的基本性质.
- 第 2 章定义了线性相关、张成、基、维数, 并给出了有限维向量空间的基础理论.
- 第 3 章引入了线性映射. 这一章的主要结果是, 线性映射的零空间的维数加上值域的维数等于定义域的维数.
- 第 4 章给出了多项式的部分理论, 这是理解线性算子所必需的. 如果在课堂上把本章的证明都讲一遍 (本章没有线性代数内容), 那么线性代数的某些重要内容可能就没时间讲了. 学生应该已经熟悉本章关于多项式的这些定理, 因此可以要求他们只看结果的陈述而不看证明. 当然, 好奇的学生还是会看其中的一些证明的, 这也正是本书包含这些证明的原因.
- 第 5 章引入了本征向量, 这源自将线性算子限制到更小

的子空间上来研究的思想. 本章的精彩之处是复向量空间上本征值存在性的简洁证明. 我们还利用这个结论, 证明了复向量空间上的线性算子关于某个基有上三角矩阵. 用类似的方法, 证明了实向量空间上的线性算子都具有 1 维或 2 维的不变子空间, 并利用此结果, 证明了奇数维实向量空间上的线性算子都有本征值. 我们的这些证明都不需要定义行列式和特征多项式.

- 第 6 章定义了内积空间, 阐述了它们的基本性质并介绍了一些标准工具, 如规范正交基、格拉姆–施密特正交化过程以及伴随. 本章还介绍了如何利用正交投影来解某些极小化问题.
- 第 7 章的精彩之处是谱定理, 它刻画了本征向量可以组成规范正交基的线性算子. 有了前几章的工作, 本章的证明都特别简单. 这一章还讨论了正定算子、线性等距同构、极分解以及奇异值分解.
- 第 8 章引入了极小多项式、特征多项式以及广义本征向量，主要成果是用广义本征向量来描述复向量空间上的线性算子. 利用这个描述可以证明出几乎所有通常要使用 Jordan 形来证明的结果. 例如, 用这些工具我们证明了复向量空间上的可逆线性算子都有平方根. 本章最后证明了复向量空间上的线性算子都有 Jordan 形.
- 第 9 章的核心是实向量空间上的线性算子. 此类算子可能没有本征值, 而 2 维不变子空间弥补了这一不足, 从而可以得到与复向量空间类似的结果.
- 在第 10 章中, 我们利用特征多项式给出了迹和行列式的定义 (前面定义特征多项式时并未使用行列式). 在复向量空间上, 这些定义还可以用另一种方式来陈述: 迹是所有本征值之和, 行列式是所有本征值之积 (两种情况都计重数). 传统的处理方法是利用行列式来证明本征值的存在性, 这不可能得到这些好记的定义. 我们现在的处理方法也使得行列式的一些标准定理变得更加清楚. 我们利用极分解和自伴算子的刻画导出了多重积分的换元公式, 这就使得行列式在其中的出现看起来非常自然.

通过取 $\mathbf{F}$ 表示实数域或复数域, 本书经常同时发展实向量空间和复向量空间上的线性代数理论. 也可以采用抽象的域, 但

这只会增加抽象性而不会导出线性代数的任何新内容. 只考虑实数和复数的另一个理由是, 可视多项式为真正的函数, 而不必像在有限域上那样把多项式看作更形式化的对象. 最后还有一点, 即使理论的开始部分可以在任意域上展开, 但是内积空间还是会把我们的讨论拉回到实向量空间和复向量空间.

即便是这么薄的一本书, 你也不要指望能把所有内容都讲完. 一学期讲完前八章就已经是一个雄心勃勃的目标了. 如果一定要讲到第 10 章, 我建议快速讲完第 1 章、第 2 章和第 4 章 (学生可能在以前的课程中学过这些内容), 并跳过第 9 章 (在这样的情况下, 你就只能在复向量空间上讨论迹和行列式).

提高学生理解和熟练运用线性代数知识的能力要比讲授任何一套特殊的定理都更重要. 数学只能做着学, 好在线性代数有很多好的作业题. 在教这门课程时, 我通常每次课留两三道习题作为作业, 要求到下次课时交. 讲解作业大概要占用一节课的三分之一甚至一半的时间.

有一份包含全部练习题答案的手册可供 (免费) 使用, 但只提供给使用本书作为教材的教师. 教师可以给我发 E-mail 索取该手册 (也可以和 Springer 出版社联系).

请到我的网站上查看本书的勘误表 (我希望它是空的或者几乎是空的) 和其他信息.

如果你能告知本书中的错误, 哪怕是很小的错误, 我都会十分感激. 欢迎为本书的改进提出建议, 哪怕是细微的改进. 请随时和我联系.

祝你教学愉快!

Sheldon Axler
美国旧金山州立大学数学系
美国旧金山, CA 94132
E-mail: axler@math.sfsu.edu

致 学 生

你可能要学习第二门线性代数课程. 在你第一次接触线性代数时, 学习的重点可能是欧氏空间和矩阵. 而本书关注的是向量空间和线性映射. 这些术语以后会定义的, 所以即使你现在不了解这些术语的含义也没关系. 本书从线性代数的初步知识讲起, 不需要线性代数方面的任何基础. 关键是你要潜心于严谨的数学, 尤其要深入地理解定义、定理、证明.

别指望读数学书能像读小说一样. 要是你不到一小时就读完一页的话, 那就可能读得太快了. 当遇到 “你应该验证” 这样的话时, 你的确需要自己动笔来验证一下. 有些证明步骤被省略了, 你要将其补充完整. 对每一个定义都要仔细琢磨, 用心体会. 对每一个定理都要试着举例说明为什么定理中的假设都是必要的.

请到我的网站上查看本书的勘误表 (我希望它是空的或者几乎是空的) 和其他信息.

如果你能告知本书中的错误, 哪怕是很小的错误, 我都会十分感激. 欢迎为本书的改进提出建议, 再小的建议我都欢迎.

祝你学习愉快!

Sheldon Axler
美国旧金山州立大学数学系
美国旧金山, CA 94132
E-mail: axler@math.sfsu.edu

致　谢

万分感谢在过去的两个世纪里为创建线性代数贡献智慧的数学家们. 在撰写本书的过程中，我尽量寻求阐述线性代数理论和证明其定理的最佳方式，而不去借鉴大多数教科书中所采用的标准方法和证明. 虽然我确实受到了过去所学习过的许多书籍的影响，但在本书的撰写过程中我并没有参考任何其他的书籍. 本书中大部分结果都属于公共的数学遗产. 定理的某个特殊情况可能在古代 (对线性代数来说这指的是 19 世纪) 被首次证明，现在我们看到的定理是众多数学家经过数十年不断地加强和完善才得到的. 但要一一列举每一位数学家的确切贡献却是一项很困难的任务，本书也就没有逐一列出. 无论如何，读者都不要把本书中的任何定理当成我的原创.

本书在很多人的帮助下才得以完善. 感谢下列人员提出宝贵建议和指正：William Arveson(建议了定理 5.13 的证明), Marilyn Brouwer, William Brown, Robert Burckel, Paul Cohn, James Dudziak, David Feldman (建议了引理 8.40 的证明), Pamela Gorkin, Aram Harrow, Pan Fong Ho, Dan Kalman, Robert Kantrowitz, Ramana Kappagantu, Mizan Khan, Mikael Lindström, Jacob Plotkin, Elena Poletaeva, Mihaela Poplicher, Richard Potter, Wade Ramey, Marian Robbins, Jonathan Rosenberg, Joan Stamm, Thomas Starbird, Jay Valanju, Thomas von Foerster.

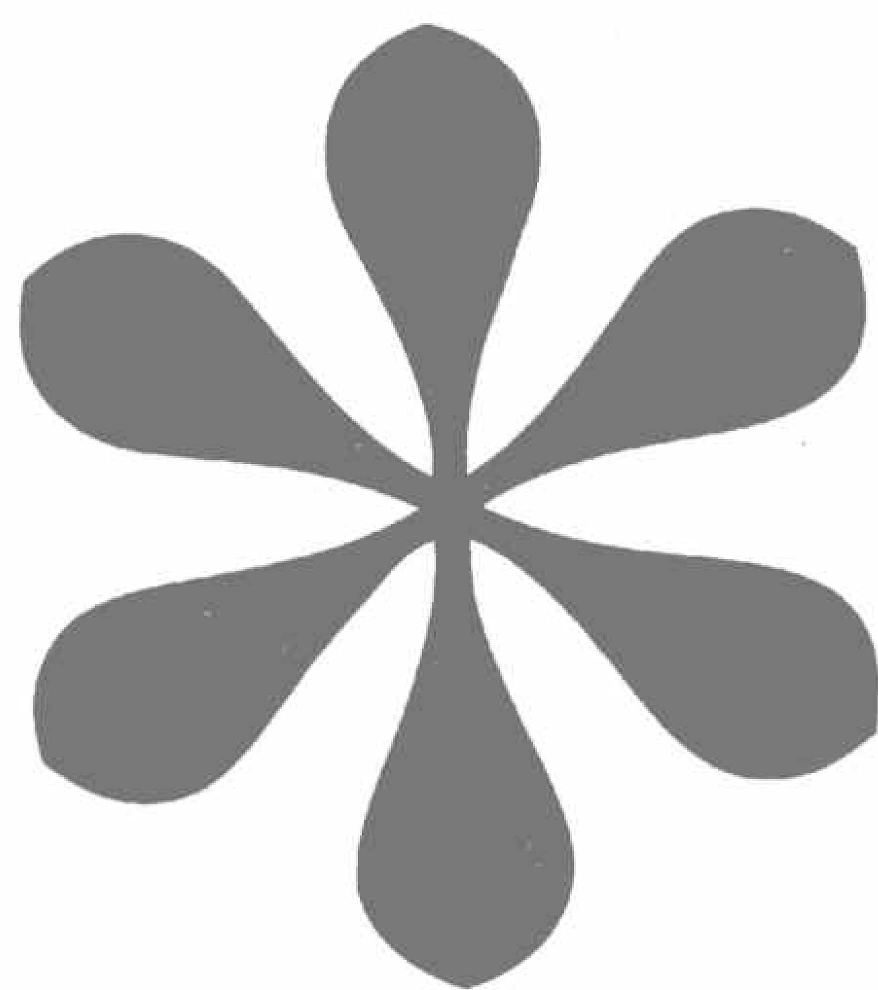
最后，感谢 Springer 出版社在我需要时所给予的帮助，并允许我最终决定本书的内容和版式.

目　录

第1章　向 量 空 间

线性代数主要研究有限维向量空间上的线性映射. 这些术语的含义我们以后会搞清楚的. 本章将给出向量空间的定义, 并讨论向量空间的基本性质.

在某些数学领域, 包括线性代数, 如果在研究实数的同时也研究复数, 就会得到更好的定理, 而且理解也会更深刻. 因此, 我们先介绍复数及其基本性质.



§1.1 复 数

你应该已经熟悉实数集 $\mathbf{R}$ 的基本性质. 复数的发明使得我们可以对负数取平方根. 关键思想是假定 -1 有平方根, 记为 i, 并按照通常的算术法则对 i 进行运算. 形式上, 一个**复数** **(complex number)** 就是一个有序的数对 (a, b), 其中 $a, b \in \mathbf{R}$, 但是我们把它写成 $a + bi$. 把所有复数构成的集合记为 $\mathbf{C}$:

瑞士数学家欧拉 (Leonhard Euler) 于 1777 年最先使用符号 i 来表示 $\sqrt{-1}$.

$$\mathbf{C} = \{a + bi : a, b \in \mathbf{R}\}.$$

若 $a \in \mathbf{R}$, 则将 $a + 0i$ 和 a 看成是一样的. 于是可以将 $\mathbf{R}$ 看作 $\mathbf{C}$ 的一个子集.

$\mathbf{C}$ 上的加法和乘法定义如下:

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i,$$

其中 $a, b, c, d \in \mathbf{R}$. 在这样的乘法下, 你应该验证 $i^2 = -1$. 不要去背上面的复数乘法公式, 只要记住 $i^2 = -1$ 并利用通常的运算法则就可以推导出这个公式.

利用熟知的实数的性质, 你应该验证, $\mathbf{C}$ 上的加法和乘法满足以下性质:

交换性 (commutativity)

对所有 $w, z \in \mathbf{C}$, 都有 $w + z = z + w$, $wz = zw$;

结合性 (associativity)

对所有 $z_1, z_2, z_3 \in \mathbf{C}$, 都有

$(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$, $(z_1 z_2) z_3 = z_1 (z_2 z_3)$;

单位元 (identities)

对所有 $z \in \mathbf{C}$, 都有 $z + 0 = z$, $z1 = z$;

加法逆 (additive inverse)

对每个 $z \in \mathbf{C}$, 都存在唯一一个 $w \in \mathbf{C}$, 使得 $z + w = 0$;

乘法逆 (multiplicative inverse)

对每个 $z \in \mathbf{C}$, $z \neq 0$, 都存在唯一一个 $w \in \mathbf{C}$, 使得 $zw = 1$;

分配性质 (distributive property)

对所有 $\lambda, w, z \in \mathbf{C}$, 都有 $\lambda(w+z) = \lambda w + \lambda z$.

对 $z \in \mathbf{C}$, 以 $-z$ 表示 z 的加法逆. 因而, $-z$ 是使得

$$z + (-z) = 0$$

的唯一复数. $\mathbf{C}$ 上的减法定义为

$$w - z = w + (-z),$$

其中 $w, z \in \mathbf{C}$.

对 $z \in \mathbf{C}$, $z \neq 0$, 以 $1/z$ 表示 z 的乘法逆. 因而, $1/z$ 是使得

$$z(1/z) = 1$$

的唯一复数. $\mathbf{C}$ 上的除法定义为

$$w/z = w(1/z),$$

其中 $w, z \in \mathbf{C}$, $z \neq 0$.

为了使给出的定义和证明的定理对于实数和复数都适用, 我们将采用如下记号:

本节中 $\mathbf{F}$ 总表示 $\mathbf{R}$ 或 $\mathbf{C}$.

选用字母 $\mathbf{F}$ 是因为 $\mathbf{R}$ 和 $\mathbf{C}$ 都是所谓**域** (field) 的例子. 本书并不讨论 $\mathbf{R}$ 和 $\mathbf{C}$ 之外的域. 线性代数中很多对 $\mathbf{R}$ 和 $\mathbf{C}$ 成立的定义、定理、证明对任意域都照样成立.

于是, 如果我们在 $\mathbf{F}$ 上证明了一个定理, 那么将其中的 $\mathbf{F}$ 换成 $\mathbf{R}$ 或 $\mathbf{C}$, 定理照样成立. $\mathbf{F}$ 中的元素称为**标量** (**scalar**). “标量”就是数的意思. 通常用标量来强调一个对象是数, 而不是向量 (马上就给出向量的定义).

对 $z \in \mathbf{F}$ 及正整数 m, 我们把 z^m 定义为 m 个 z 的乘积:

$$z^m = \underbrace{z \cdots\cdots z}_{m\text{个}}.$$

显然, 对所有 $w, z \in \mathbf{F}$ 及正整数 m, n, 都有 $(z^m)^n = z^{mn}$, $(wz)^m = w^m z^m$.

§1.2 向量空间的定义

在给出向量空间的定义之前先来看两个重要的例子. 向量空间 $\mathbf{R}^2$ 可以看作一个平面, 它由所有有序实数对构成:

$$\mathbf{R}^2 = \{(x, y) : x, y \in \mathbf{R}\}.$$

向量空间 $\mathbf{R}^3$ 可以看作通常的空间, 它由所有有序三元实数组构成:

$$\mathbf{R}^3 = \{(x, y, z) : x, y, z \in \mathbf{R}\}.$$

为了把 $\mathbf{R}^2$ 和 $\mathbf{R}^3$ 推广到更高维, 需要先讨论组的概念. 设 n 是一个非负整数. **长度 (length)** 为 n 的**组 (list)** 是按序排列、用逗号隔开并且两端用括弧括起来的 n 个对象 (可以是数、其他组或者更抽象的东西). 一个长度为 n 的组具有如下形式:

很多数学家称长度为 n 的组为 n **元组 (tuple)**.

$$(x_1, \cdots, x_n).$$

于是, 长度为 2 的组是有序对, 而长度为 3 的组是有序 3 元组. 对于 $j \in \{1, \cdots, n\}$, 称 x_j 是上述组的第 j 个**坐标 (coordinate)**. 因此, x_1 称为第一个坐标, x_2 称为第二个坐标, 依此类推.

有时候我们只说**组 (list)** 而不指明其长度. 但要记住, 根据定义, 每个组的长度都是有限的, 这个长度是一个非负整数. 因此, 形如

$$(x_1, x_2, \cdots)$$

的对象不是组, 或许可以称其具有无限长度. 长度为 0 的组形如: (); 将其视为组, 可使一些定理避免平凡的例外.

两个组相等当且仅当它们长度相同并且对应的坐标相等. 也就是说, $(x_1, \cdots, x_m)$ 等于 $(y_1, \cdots, y_n)$ 当且仅当 $m = n$ 并且 $x_1 = y_1, \cdots, x_m = y_m$.

组与集合有两点不同: 组是有顺序并且允许重复的, 而对于集合来说, 顺序和重复都无关紧要. 例如, 组 $(3, 5)$ 和 $(5, 3)$ 是不相等的, 但是集合 $\{3, 5\}$ 和 $\{5, 3\}$ 是相等的. 组 $(4, 4)$ 和 $(4, 4, 4)$ 是不相等的 (它们的长度不同), 而集合 $\{4, 4\}$ 和 $\{4, 4, 4\}$ 都等于集合 $\{4\}$.

要定义 $\mathbf{R}^2$ 和 $\mathbf{R}^3$ 的高维相似物, 只要用 $\mathbf{F}$ (等于 $\mathbf{R}$ 或 $\mathbf{C}$) 代替 $\mathbf{R}$, 并且用任意正整数代替 2 或 3 即可. 特别地, 在本节的其余部分, 我们将固定一个正整数 n. 定义 $\mathbf{F}^n$ 为 $\mathbf{F}$ 中元素组成的长度为 n 的组的集合:

$$\mathbf{F}^n = \{(x_1, \cdots, x_n) : x_j \in \mathbf{F}, j = 1, \cdots, n\}.$$

例如, 若 $\mathbf{F} = \mathbf{R}$ 并且 n 等于 2 或 3, 则 $\mathbf{F}^n$ 的这个定义与前面 $\mathbf{R}^2$ 和 $\mathbf{R}^3$ 的定义是一致的. 又如, $\mathbf{C}^4$ 是所有含 4 个复数的组所构成的集合:

$$\mathbf{C}^4 = \{(z_1, z_2, z_3, z_4) : z_1, z_2, z_3, z_4 \in \mathbf{C}\}.$$

若 $n \geqslant 4$, 则将 $\mathbf{R}^n$ 想象成一个物理对象并非易事. 如果我们讨论复数, 也会有同样的问题: $\mathbf{C}^1$ 可以看成一个平面, 但是对于 $n \geqslant 2$, 人类的大脑不能产生 $\mathbf{C}^n$ 的几何模型. 不过即使 n 很大, 我们仍可以在 $\mathbf{F}^n$ 上进行代数运算, 而且就像在 $\mathbf{R}^2$ 或者 $\mathbf{R}^3$ 上一样地容易. 例如, $\mathbf{F}^n$ 上的加法可以通过相应坐标相加来定义:

关于生活在 $\mathbf{R}^2$ 中的生物如何感知 $\mathbf{R}^3$ 的有趣描述, 可读一读 Edwin A. Abbott 所著的 ***Flatland: A Romance of Many Dimensions***. 这部 1884 年出版的小说可以帮助像我们一样生活在三维空间中的生物想象四维和更高维的物理空间.

1.1 $$(x_1, \cdots, x_n) + (y_1, \cdots, y_n) = (x_1 + y_1, \cdots, x_n + y_n).$$

如果我们采用单个字母来表示含有 n 个数的组, 而不明确地写出每一个坐标, 那么 $\mathbf{F}^n$ 上的数学通常会变得更简洁. 因此, $\mathbf{F}^n$ 上的加法交换性可以表述成: 对所有 $x, y \in \mathbf{F}^n$ 都有

$$x + y = y + x,$$

而不必更繁琐地写成 (即使在证明交换性时需要下面这个公式): 对所有 $x_1, \cdots, x_n, y_1, \cdots, y_n \in \mathbf{F}$ 都有

$$(x_1, \cdots, x_n) + (y_1, \cdots, y_n) = (y_1, \cdots, y_n) + (x_1, \cdots, x_n).$$

如果用单个字母来表示 $\mathbf{F}^n$ 中的一个元素, 那么在必须列出坐标时, 通常用带有适当下标的同一字母来表示其中的坐标. 例如, 若 $x \in \mathbf{F}^n$, 则令 x 等于 $(x_1, \cdots, x_n)$ 就是很好的记法. 如果可能的话, 只讨论 x 而省略具体的坐标会更好.

用 0 表示长度为 n 且所有坐标都为 0 的组：

$$0=(0,\cdots,0).$$

注意, 我们用两种不同的方式使用符号 0 —— 在上式的左边, 0 表示一个长度为 n 的组, 而在右边, 每个 0 都表示一个数. 这种有可能引起混乱的做法实际上不会产生任何问题, 因为上下文会表明 0 指的是什么. 例如, 考虑陈述：对于 $\mathbf{F}^n$ 的加法单位元 0 有,

$$x+0=x,\quad x\in\mathbf{F}^n,$$

这里 0 必然是一个组, 因为我们从未定义过 $\mathbf{F}^n$ 中元素 (即 x) 与数 0 的和.

图形往往有助于直观. 因为我们很容易把 $\mathbf{R}^2$ 勾画在如纸或黑板这样的二维表面上, 所以我们可以通过画图来描绘这个空间. $\mathbf{R}^2$ 上的典型元素是点 $x=(x_1,x_2)$. 有时候我们不把 x 看作一个点, 而是看作一个始于原点终于 (x_1,x_2) 的箭头, 如图 1-1 所示. 当把 x 看作一个箭头时, 则称之为**向量 (vector)**.

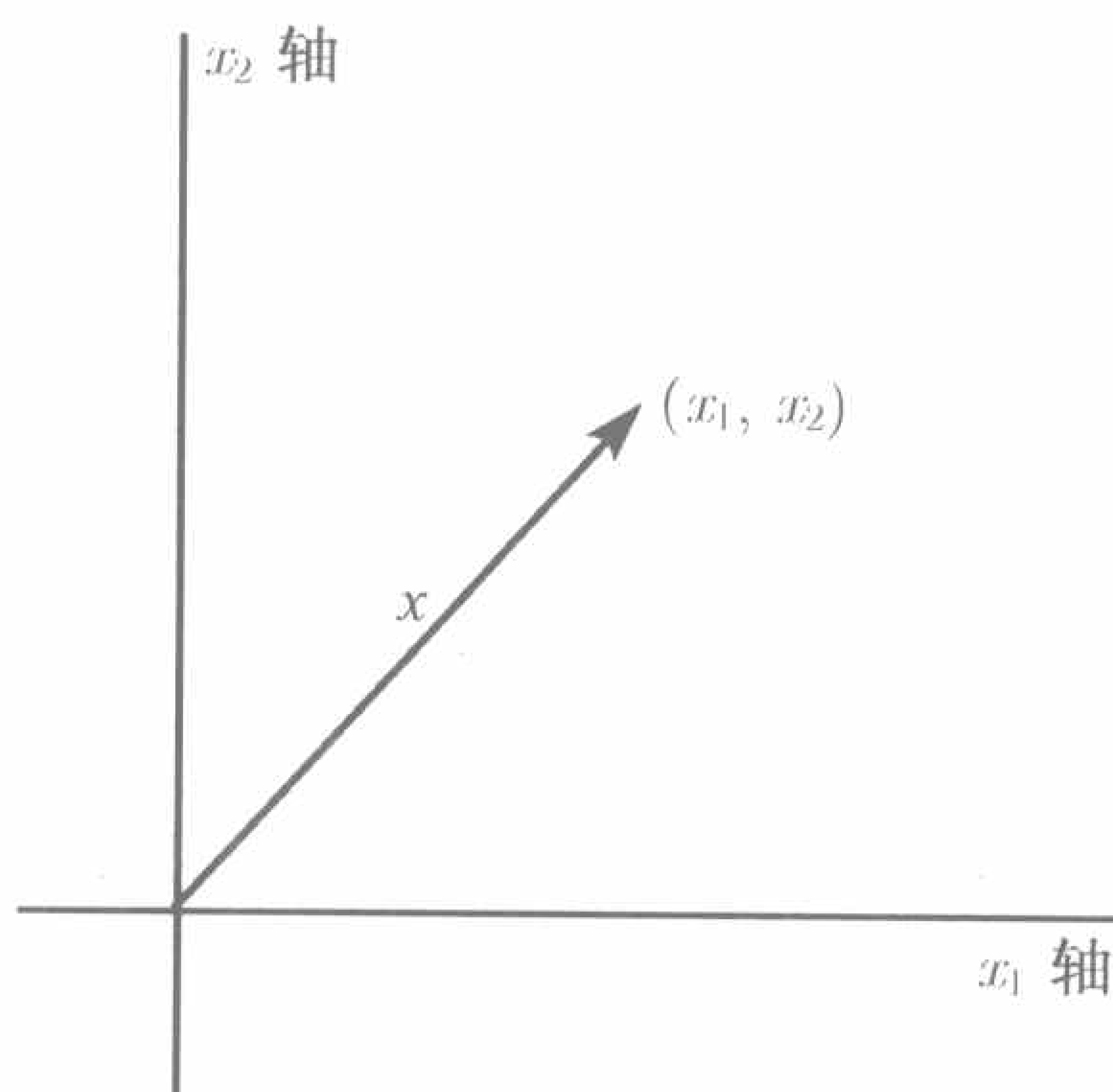


图 1-1 $\mathbf{R}^2$ 的元素可看成点或向量

这些坐标轴和具体的坐标使得图 1-1 有点乱, 省略它们而只考虑向量往往更好理解, 如图 1-2.

每当我们使用 $\mathbf{R}^2$ 中的图形或者关于点和向量的有些含糊的语言时, 要记住, 这只是为了帮助理解, 而不是要取代将要发展的真正的数学. 虽然我们画不好高维空间中的图, 但是高维空间

中的元素也像 $\mathbf{R}^2$ 中的元素一样被严格地定义. 例如, $(2, -3, 17, \pi, \sqrt{2})$ 是 $\mathbf{R}^5$ 中的元素, 而且我们偶尔还把它看作 $\mathbf{R}^5$ 中的点或者 $\mathbf{R}^5$ 中的向量, 而不用担心 $\mathbf{R}^5$ 的几何是否有物理意义.

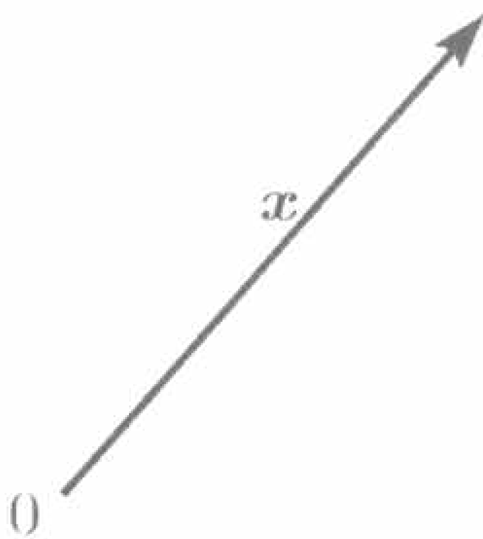


图 1-2 向量

回想一下, 我们通过相应坐标相加来定义 $\mathbf{F}^n$ 中两个元素的和, 参见 1.1. 特别地, 在 $\mathbf{R}^2$ 中, 加法有简单的几何解释. 假设我们要把 $\mathbf{R}^2$ 中的两个向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 加起来, 如图 1-3 左侧所示. 把向量 $\boldsymbol{y}$ 平行移动使其始点与向量 $\boldsymbol{x}$ 的终点重合. 那么, $\boldsymbol{x}+\boldsymbol{y}$ 就是以 $\boldsymbol{x}$ 的始点为始点, 以平移后的 $\boldsymbol{y}$ 的终点为终点的向量, 如图 1-3 右侧所示.

经济学中的数学模型通常有上千个变量, 比如说 $x_1, \cdots, x_{5000}$, 这就意味着必须在 $\mathbf{R}^{5000}$ 中进行运算. 这样的空间不能通过几何手段来处理, 但是代数方法却可以很好地发挥作用. 这就是我们的课程称为线性**代数**的原因.

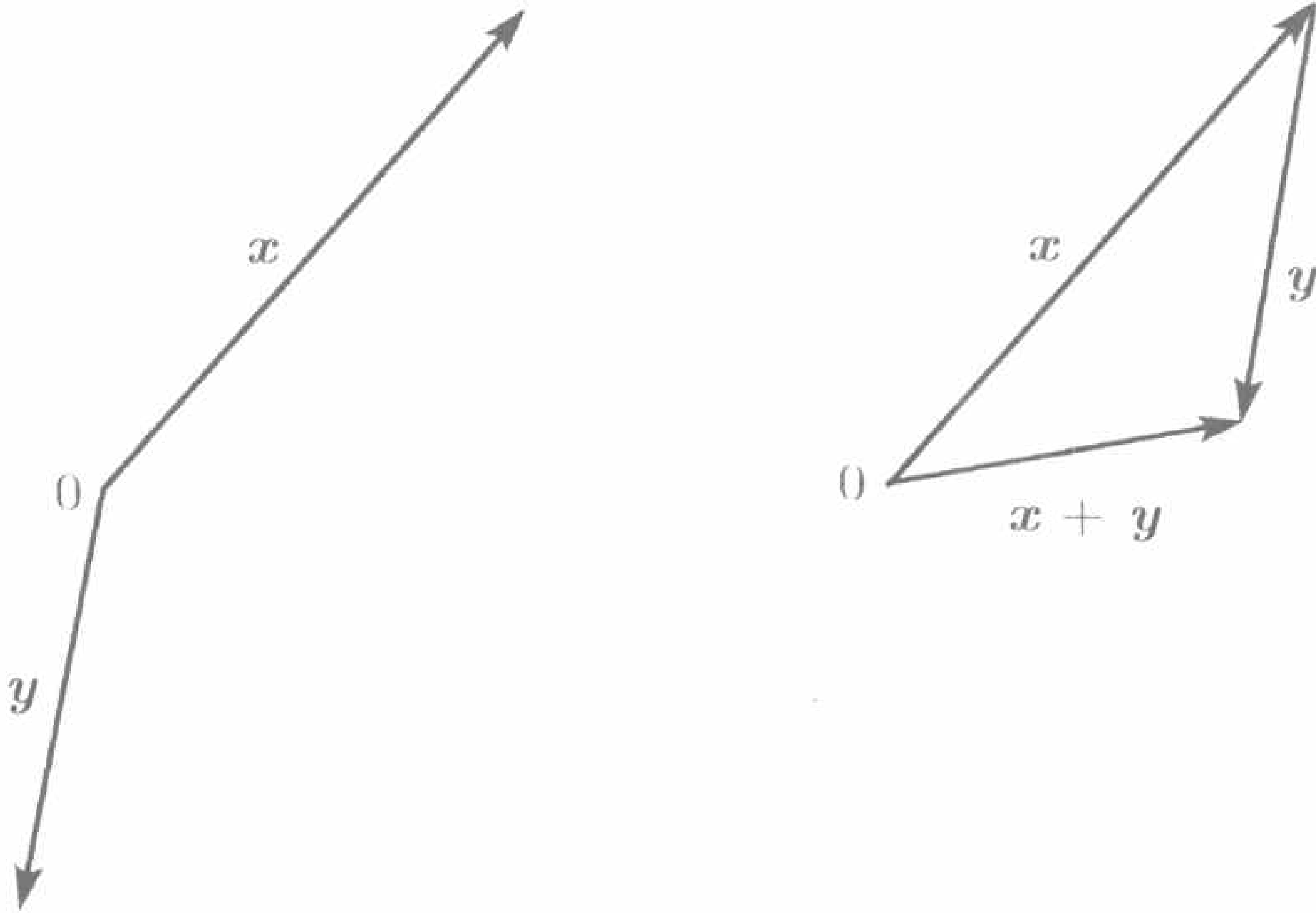


图 1-3 两个向量的和

图 1-3 中对于向量 $\boldsymbol{y}$ 的处理阐述了一个普遍原理: 在把 $\mathbf{R}^2$ 中的向量看作箭头时, 可以把箭头平行移动 (不改变它的长度和方向), 并且仍将其视为同一向量.

在讨论了 $\mathbf{F}^n$ 上的加法之后, 现在来讨论乘法. 我们或许可以用类似的方法来定义 $\mathbf{F}^n$ 上的乘法, 即通过 $\mathbf{F}^n$ 中的两个元素的对应坐标相乘来得到 $\mathbf{F}^n$ 中的另一个元素. 但经验证明, 这样

的定义对我们是没有用的. 另一种乘法, 称为标量乘法, 将在线性代数中占有重要地位. 具体地, 我们需要定义 $\mathbf{F}$ 中元素与 $\mathbf{F}^n$ 中元素的乘法. 通过乘每个坐标, 可以给出一个显然的定义:

$$a(x_1,\cdots,x_n)=(ax_1,\cdots,ax_n);$$

其中 $a\in\mathbf{F}$, $(x_1,\cdots,x_n)\in\mathbf{F}^n$.

$\mathbf{R}^2$ 中的标量乘法有很好的几何解释. 如果 a 是一个正数, $\boldsymbol{x}$ 是 $\mathbf{R}^2$ 中的一个向量, 则向量 $a\boldsymbol{x}$ 的方向与 $\boldsymbol{x}$ 相同, 而其长度为 $\boldsymbol{x}$ 的长度的 a 倍. 也就是说, 为了得到 $a\boldsymbol{x}$, 只需将 $\boldsymbol{x}$ 收缩或者伸长 a 倍, 这取决于 $a<1$ 或者 $a>1$, 如图 1-4 所示.

在标量乘法中, 我们把一个标量和一个向量相乘, 得到一个向量. 你也许熟悉 $\mathbf{R}^2$ 或者 $\mathbf{R}^3$ 中的点积, 它把两个向量相乘而得到一个标量. 当我们在第 6 章学习内积时, 点积的推广会变得很重要. 也许你还熟悉 $\mathbf{R}^3$ 中的叉积, 它将两个向量相乘而得到另一个向量. 这种乘法在高维情形不存在有用的推广.

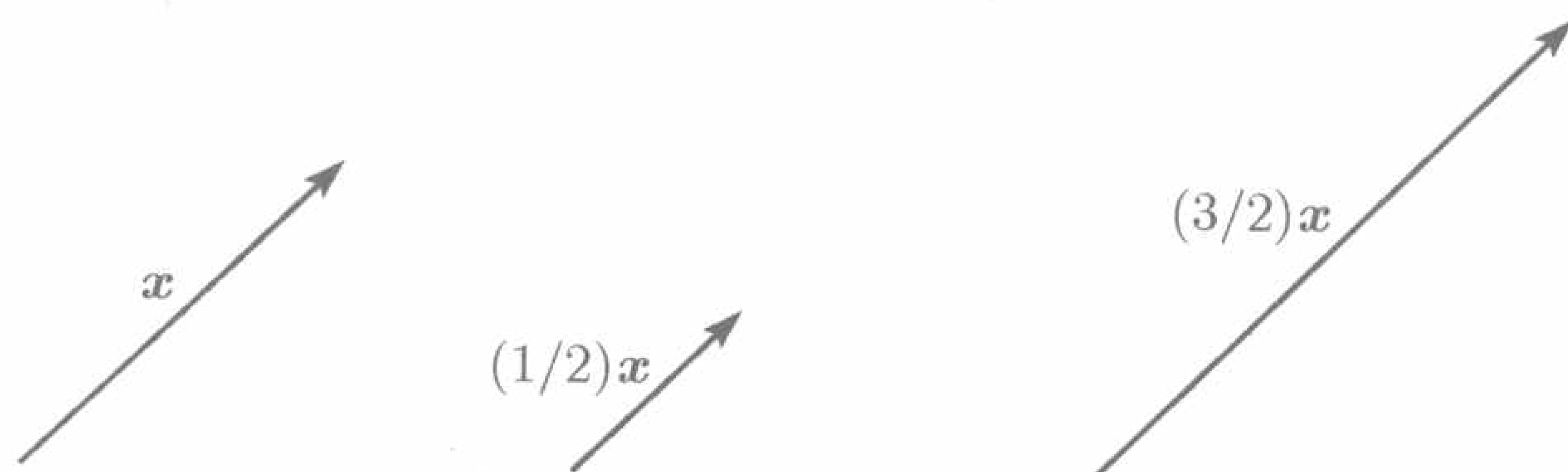


图 1-4 乘以正标量

如果 a 是一个负数, $\boldsymbol{x}$ 是 $\mathbf{R}^2$ 中的一个向量, 则向量 $a\boldsymbol{x}$ 的方向与 $\boldsymbol{x}$ 相反, 而其长度为 $\boldsymbol{x}$ 的长度的 $|a|$ 倍, 如图 1-5 所示.

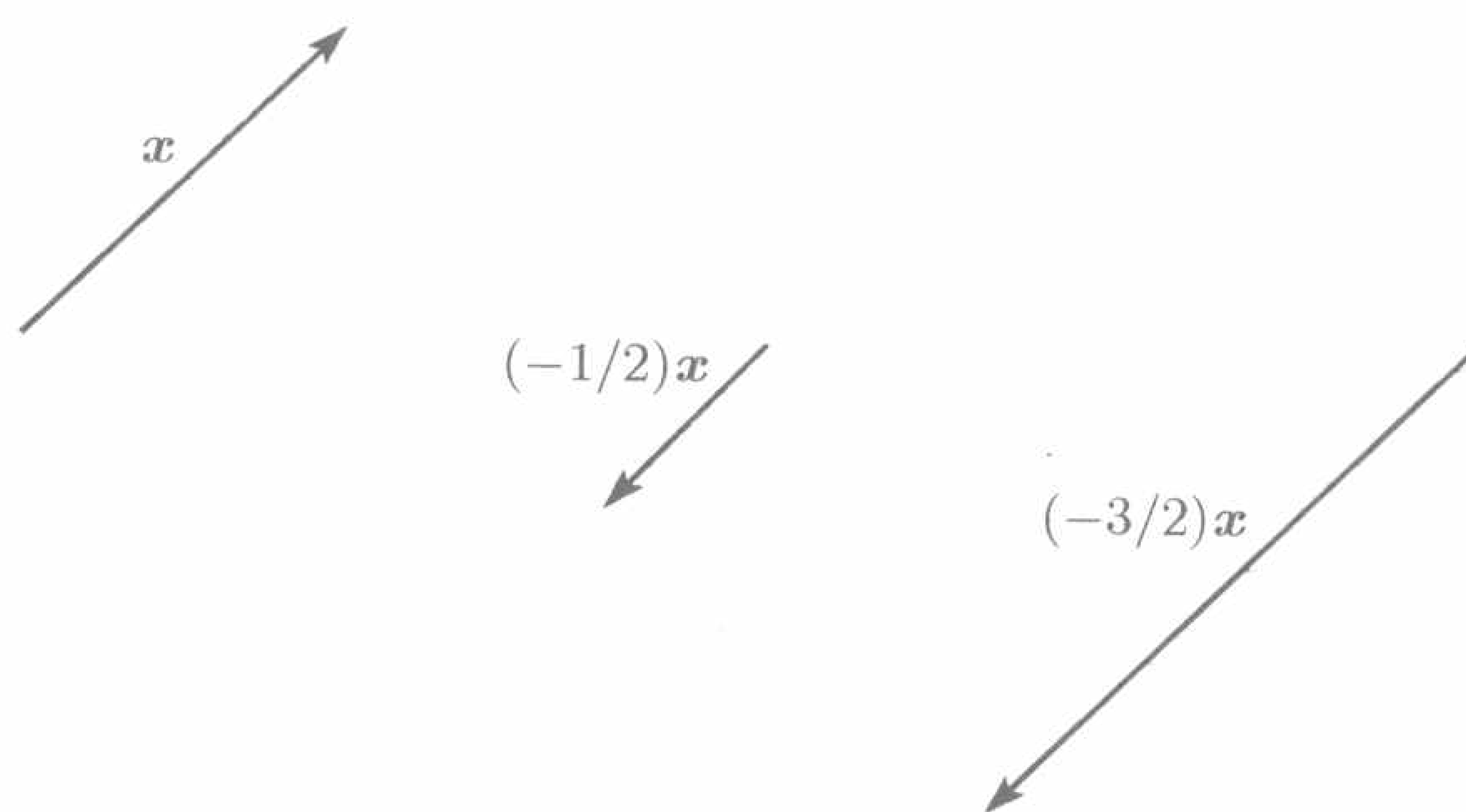


图 1-5 乘以负标量

定义向量空间的动机来自 $\mathbf{F}^n$ 上加法和标量乘法所具有的重要性质. 具体地, $\mathbf{F}^n$ 上的加法具有交换性和结合性, 并且有单位元, 即 0. 每个元素都有加法逆. $\mathbf{F}^n$ 中的标量乘法具有结合

性, 并且 1 就像乘法单位元一样. 最后, $\mathbf{F}^n$ 上的加法和标量乘法通过分配性质联系到一起.

我们把向量空间定义为带有加法和标量乘法的集合 V, 使得加法和标量乘法具有上一段所述的性质. V 上的**加法** (**addition**) 指的是一个函数, 它把每一对 $\boldsymbol{u},\boldsymbol{v}\in V$ 都对应到 V 的一个元素 $\boldsymbol{u}+\boldsymbol{v}$. V 上的**标量乘法** (**scalar multiplication**) 也是一个函数, 它把任意 $a\in\mathbf{F}$, $\boldsymbol{v}\in V$ 对应到 V 的一个元素 $a\boldsymbol{v}\in V$.

现在我们就可以给向量空间一个正式的定义. **向量空间** (**vector space**) 就是带有加法和标量乘法的集合 V, 使得下列性质成立:

交换性 (commutativity)

对所有 $\boldsymbol{u},\boldsymbol{v}\in V$, 都有 $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$;

结合性 (associativity)

对所有 $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in V$, $a,b\in\mathbf{F}$, 都有

$(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$, $(ab)\boldsymbol{v}=a(b\boldsymbol{v})$;

加法单位元 (additive identity)

存在一个元素 $\mathbf{0}\in V$, 使得对所有 $\boldsymbol{v}\in V$ 都有 $\boldsymbol{v}+\mathbf{0}=\boldsymbol{v}$;

加法逆 (additive inverse)

对每个 $\boldsymbol{v}\in V$, 都存在 $\boldsymbol{w}\in V$ 使得 $\boldsymbol{v}+\boldsymbol{w}=0$;

乘法单位元 (multiplicative identity)

对所有 $\boldsymbol{v}\in V$, 都有 $1\boldsymbol{v}=\boldsymbol{v}$;

分配性质 (distributive properties)

对所有 $a,b\in\mathbf{F}$, $\boldsymbol{u},\boldsymbol{v}\in V$, 都有

$a(\boldsymbol{u}+\boldsymbol{v})=a\boldsymbol{u}+a\boldsymbol{v}$, $(a+b)\boldsymbol{u}=a\boldsymbol{u}+b\boldsymbol{u}$.

向量空间中的标量乘法依赖于 $\mathbf{F}$. 因此, 在需要确切指明时, 我们会说 V 是 $\mathbf{F}$ 上的向量空间, 而不是简单地说 V 是向量空间. 例如, $\mathbf{R}^n$ 是 $\mathbf{R}$ 上的向量空间, $\mathbf{C}^n$ 是 $\mathbf{C}$ 上的向量空间. $\mathbf{R}$ 上的向量空间通常称为**实向量空间** (**real vector space**), $\mathbf{C}$ 上的向量空间通常称为**复向量空间** (**complex vector space**). $\mathbf{F}$ 的选取往往在上下文中是很明显的或者是无关紧要的, 所以, 我们通常都假设 $\mathbf{F}$ 是自明的.

向量空间的元素称为**向量** (**vector**) 或**点** (**point**). 这种几何语言有助于直观.

毫不意外, 你应该验证 $\mathbf{F}^n$ 是 $\mathbf{F}$ 上的向量空间. 当然, 这个例子为我们定义向量空间提供了动机.

最简单的向量空间只含有一个点, 即 $\{\mathbf{0}\}$ 是向量空间, 但这是一个没多大意思的向量空间.

又如, 考虑 $\mathbf{F}^\infty$, 它是由 $\mathbf{F}$ 中元素的所有序列构成的集合:

$$\mathbf{F}^\infty = \{(x_1, x_2, \cdots) : x_j \in \mathbf{F}, \quad j = 1, 2, \cdots\}.$$

$\mathbf{F}^\infty$ 中的加法和标量乘法的定义也和我们所料想的一样:

$$(x_1, x_2, \cdots) + (y_1, y_2, \cdots) = (x_1 + y_1, x_2 + y_2, \cdots),$$

$$a(x_1, x_2, \cdots) = (ax_1, ax_2, \cdots).$$

你应该验证, 在此定义下, $\mathbf{F}^\infty$ 成为 $\mathbf{F}$ 上的向量空间, 其中的加法单位元是每个元素都为 0 的序列.

下一个向量空间的例子是关于多项式的. 一个函数 $\boldsymbol{p} : \mathbf{F} \to \mathbf{F}$ 称为系数在 $\mathbf{F}$ 中的**多项式 (polynomial)**, 如果存在 $a_0, \cdots, a_m \in \mathbf{F}$ 使得

$$\boldsymbol{p}(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m, \quad z \in \mathbf{F}.$$

虽然 $\mathbf{F}^n$ 是向量空间的一个至关紧要的例子, 但并不是所有的向量空间都是由组构成的. 例如, $\mathcal{P}(\mathbf{F})$ 的元素是 $\mathbf{F}$ 上的函数, 而不是组. 一般来说, 向量空间是一个抽象的对象, 其中的元素可能是组、函数或稀奇古怪的对象.

我们定义 $\mathcal{P}(\mathbf{F})$ 为系数在 $\mathbf{F}$ 中的所有多项式构成的集合. $\mathcal{P}(\mathbf{F})$ 上加法的定义正如你所料: 若 $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{P}(\mathbf{F})$, 则 $\boldsymbol{p} + \boldsymbol{q}$ 就是如下定义的多项式

$$(\boldsymbol{p} + \boldsymbol{q})(z) = \boldsymbol{p}(z) + \boldsymbol{q}(z), \quad z \in \mathbf{F}.$$

例如, 若 $\boldsymbol{p}$ 为多项式 $\boldsymbol{p}(z) = 2z + z^3$, $\boldsymbol{q}$ 为多项式 $\boldsymbol{q}(z) = 7 + 4z$, 则 $\boldsymbol{p} + \boldsymbol{q}$ 为多项式 $(\boldsymbol{p} + \boldsymbol{q})(z) = 7 + 6z + z^3$. $\mathcal{P}(\mathbf{F})$ 上的标量乘法的定义也是显然的: 若 $a \in \mathbf{F}$, $\boldsymbol{p} \in \mathcal{P}(\mathbf{F})$, 则 $a\boldsymbol{p}$ 就是如下定义的多项式

$$(a\boldsymbol{p})(z) = a\boldsymbol{p}(z), \quad z \in \mathbf{F}.$$

对于这样定义的加法和标量乘法, 你应该验证, $\mathcal{P}(\mathbf{F})$是向量空间, 其中的加法单位元是所有系数都为 0 的多项式.

我们很快就会看到更多的向量空间的例子, 但是我们需要先来描述向量空间的一些基本性质.

§1.3 向量空间的性质

向量空间的定义要求向量空间具有加法单位元. 下面的命题说明这个单位元是唯一的.

1.2 命题: *向量空间有唯一的加法单位元.*

证明: 设 $\mathbf{0}$ 和 $\mathbf{0}'$ 都是向量空间 V 的加法单位元, 则

$$\mathbf{0}' = \mathbf{0}' + \mathbf{0} = \mathbf{0},$$

其中第一个等式成立是因为 $\mathbf{0}$ 是加法单位元, 第二个等式成立是因为 $\mathbf{0}'$ 是加法单位元. 因此 $\mathbf{0} = \mathbf{0}'$, 这就证明了 V 中只有一个加法单位元. ■

符号 ■ 的意思是“证明结束”.

向量空间的每个元素 v 都有加法逆, 即向量空间中使得 $v + w = \mathbf{0}$ 的元素 w. 下一个命题表明向量空间中的每个元素都只有一个加法逆.

1.3 命题: *向量空间中的每个元素都有唯一的加法逆.*

证明: 设 V 是向量空间, $v \in V$, 并且 w 和 w' 都是 v 的加法逆, 那么

$$w = w + \mathbf{0} = w + (v + w') = (w + v) + w' = \mathbf{0} + w' = w'.$$

因此 $w = w'$. ■

由于加法逆是唯一的, 因此可以令 $-v$ 表示向量 v 的加法逆. 定义 $w - v$ 为 $w + (-v)$.

本书中几乎所有的结论都会涉及向量空间. 为了避免经常重复像“设 V 是向量空间”这样的陈述, 我们现在作出必要的声明, 从此便可以一劳永逸:

在本书的其余部分, 总设 V 是 $\mathbf{F}$ 上的向量空间.

因为有结合性, 所以在处理向量空间中两个以上元素的加法时可以省略括号. 例如, 可以写成 $u + v + w$, 而无需括号, 因

为此表达式的两种可能的解释 $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}$ 和 $\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$ 是相等的. 我们首先在下一个证明中采用这种不加括号的自然约定. 在下一个命题中, 等式左边的 0 表示标量 (数 $0 \in \mathbf{F}$), 等式右边的 $\mathbf{0}$ 表示向量 (V 中的加法单位元).

注意, 1.4 和 1.5 是关于标量乘法和 V 的加法单位元的论断. 在向量空间的定义中, 唯一将标量乘法和向量的加法联系在一起的是分配性质. 因此在证明中必然要用到分配性质.

1.4 命题: 对每个 $\boldsymbol{v} \in V$ 都有 $0\boldsymbol{v} = \mathbf{0}$.

证明: 对 $\boldsymbol{v} \in V$, 有

$$0\boldsymbol{v} = (0+0)\boldsymbol{v} = 0\boldsymbol{v} + 0\boldsymbol{v}.$$

在上面等式的两端都加上 $0\boldsymbol{v}$ 的加法逆, 可得 $\mathbf{0} = 0\boldsymbol{v}$. ■

在下一个命题中, $\mathbf{0}$ 表示 V 的加法单位元. 虽然 1.4 和 1.5 的证明是相似的, 但它们并不是一回事. 确切地说, 1.4 说明标量 0 和任意向量的乘积都等于向量 $\mathbf{0}$, 而 1.5 说明任意标量与向量 $\mathbf{0}$ 的乘积都等于向量 $\mathbf{0}$.

1.5 命题: 对每个 $a \in \mathbf{F}$ 都有 $a\mathbf{0} = \mathbf{0}$.

证明: 对 $a \in \mathbf{F}$, 有

$$a\mathbf{0} = a(\mathbf{0}+\mathbf{0}) = a\mathbf{0} + a\mathbf{0}.$$

在上面等式的两端都加上 $a\mathbf{0}$ 的加法逆, 可得 $\mathbf{0} = a\mathbf{0}$. ■

现在我们来证明, 若将 V 中元素与标量 -1 相乘, 则得到这个元素的加法逆.

1.6 命题: 对每个 $\boldsymbol{v} \in V$ 都有 $(-1)\boldsymbol{v} = -\boldsymbol{v}$.

证明: 对 $\boldsymbol{v} \in V$, 有

$$\boldsymbol{v} + (-1)\boldsymbol{v} = 1\boldsymbol{v} + (-1)\boldsymbol{v} = (1+(-1))\boldsymbol{v} = 0\boldsymbol{v} = \mathbf{0}.$$

这个等式说明, $(-1)\boldsymbol{v}$ 与 $\boldsymbol{v}$ 相加得 $\mathbf{0}$. 因此 $(-1)\boldsymbol{v}$ 必为 $\boldsymbol{v}$ 的加法逆. ■

§1.4 子 空 间

V 的子集 U 称为 V 的**子空间** (subspace), 如果 U (采用与 V 相同的加法和标量乘法) 也是向量空间. 例如,

$$\{(x_1, x_2, 0) : x_1, x_2 \in \mathbf{F}\}$$

是 $\mathbf{F}^3$ 的一个子空间.

有些数学家采用术语**线性子空间** (linear subspace), 意思与子空间相同.

如果 U 是 V 的子集, 要验证 U 是 V 的子空间, 只需验证 U 满足下列性质:

加法单位元 (additive identity)

$\mathbf{0} \in U$;

对加法封闭 (closed under addition)

若 $\boldsymbol{u}, \boldsymbol{v} \in U$, 则 $\boldsymbol{u} + \boldsymbol{v} \in U$;

对标量乘法封闭 (closed under scalar multiplication)

若 $a \in \mathbf{F}$, $\boldsymbol{u} \in U$, 则 $a\boldsymbol{u} \in U$.

第一个条件保证了 V 的加法单位元在 U 中, 第二个条件保证了加法在 U 上是有意义的, 第三个条件保证了标量乘法在 U 上是有意义的. 要证明 U 是向量空间, 不需要验证向量空间定义中的其他部分, 因为这些都是自然成立的. 例如, 因为加法结合性和交换性在更大的空间 V 上成立, 所以在 U 上自动成立. 又如, 若上面的第三个条件成立, 并且 $\boldsymbol{u} \in U$, 则 $-u$ (由 1.6, 它等于 $(-1)\boldsymbol{u}$) 也包含于 U, 因此 U 的每个元素在 U 中都有加法逆.

显然, $\{\mathbf{0}\}$ 是 V 的最小的子空间, V 自身是 V 的最大的子空间. 空集不是 V 的子空间, 这是因为子空间必须是向量空间, 而向量空间至少要包含一个元素, 即加法单位元.

上面的三个条件经常使我们可以快速判断一个给定的集合是不是 V 的子空间. 例如, 若 $b \in \mathbf{F}$, 则可以证明

$$\{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_3 = 5x_4 + b\}$$

是 $\mathbf{F}^4$ 的子空间当且仅当 $b = 0$. 又如, 可以证明

$$\{\boldsymbol{p} \in \mathcal{P}(\mathbf{F}) : \boldsymbol{p}(3) = 0\}$$

是 $\mathcal{P}(\mathbf{F})$ 的子空间.

$\mathbf{R}^2$ 的子空间恰好为 $\{\mathbf{0}\}$, $\mathbf{R}^2$, $\mathbf{R}^2$ 中所有过原点的直线. $\mathbf{R}^3$ 的子空间恰好为 $\{\mathbf{0}\}$, $\mathbf{R}^3$, $\mathbf{R}^3$ 中所有过原点的直线、$\mathbf{R}^3$ 中所有

过原点的平面. 容易证明这些对象确实都是子空间, 而难点在于证明它们是 $\mathbf{R}^2$ 或 $\mathbf{R}^3$ 仅有的子空间. 当我们在下一章引进更多的工具之后, 这样的工作就会变得比较容易.

§1.5 和与直和

在以后的各章中, 我们会发现向量空间的和与直和的概念都是很有用的. 现在我们就给出这些概念的定义.

在研究向量空间时, 我们感兴趣的通常只是子空间, 而不是任意子集. 子空间的并一般不是子空间(见本章习题9), 这正是我们通常只讨论子空间的和而不讨论并的原因.

设 $U_1, \cdots, U_m$ 都是 V 的子空间, 则 $U_1, \cdots, U_m$ 的**和 (sum)**, 记为 $U_1 + \cdots + U_m$, 定义为 $U_1, \cdots, U_m$ 中元素所有可能的和所构成的集合. 更确切地说,

$$U_1 + \cdots + U_m = \{\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_m : \boldsymbol{u}_1 \in U_1, \cdots, \boldsymbol{u}_m \in U_m\}.$$

你应该验证, 如果 $U_1, \cdots, U_m$ 都是 V 的子空间, 那么它们的和 $U_1 + \cdots + U_m$ 也是 V 的子空间.

我们来看子空间和的一些例子. 假设 U 是 $\mathbf{F}^3$ 中第 2 个和第 3 个坐标均为 0 的那些元素所构成的集合, W 是 $\mathbf{F}^3$ 中第 1 个和第 3 个坐标都为 0 的那些元素所构成的集合:

$$U = \{(x, 0, 0) \in \mathbf{F}^3 : x \in \mathbf{F}\}, \quad W = \{(0, y, 0) \in \mathbf{F}^3 : y \in \mathbf{F}\}.$$

在向量空间理论中, 子空间的和类似于集合论中子集的并. 给定向量空间的两个子空间, 包含它们的最小子空间是它们的和. 类似地, 给定一个集合的两个子集, 包含它们的最小子集是它们的并集.

那么你应该验证

1.7 $$U + W = \{(x, y, 0) : x, y \in \mathbf{F}\}.$$

又如, 假设 U 如上, W 是 $\mathbf{F}^3$ 中前两个坐标相等, 第 3 个坐标为 0 的那些元素所构成的集合:

$$W = \{(y, y, 0) \in \mathbf{F}^3 : y \in \mathbf{F}\}.$$

那么你应该验证: $U + W$ 也是 1.7 所给出的那个集合.

假设 $U_1, \cdots, U_m$ 都是 V 的子空间. 显然 $U_1, \cdots, U_m$ 都包含于 $U_1 + \cdots + U_m$ (为说明这一点, 考虑和 $\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_m$, 其中除一项之外的其余项均为 $\mathbf{0}$). 反之, V 中任何包含 $U_1, \cdots, U_m$ 的子空间一定都包含 $U_1 + \cdots + U_m$ (因为子空间包含其中元素的所有有限和). 于是, $U_1 + \cdots + U_m$ 是 V 中包含 $U_1, \cdots, U_m$ 的最小的子空间.

假设 $U_1, \cdots, U_m$ 是 V 的子空间, 使得 $V = U_1 + \cdots + U_m$, 则 V 中每个元素都可写成如下形式

$$\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_m,$$

其中每个 $\boldsymbol{u}_j \in U_j$. 我们对 V 中每个向量都可以唯一地表示成上述形式的情形特别感兴趣, 这种情形非常重要, 所以给它起一个特殊的名字: **直和 (direct sum)**. 具体地, 如果 V 的每个元素都可以唯一地写成 $\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_m$, 其中 $\boldsymbol{u}_j \in U_j$, 则称 V 是子空间 $U_1, \cdots, U_m$ 的直和, 记为 $V = U_1 \oplus \cdots \oplus U_m$.

符号 $\oplus$, 由一个圈和一个位于其内的加号组成, 用来表示直和, 以提醒我们正在处理一种特殊类型的子空间和——直和中的每个元都可以唯一地表示成这些给定的子空间中元素的和.

我们来看直和的一些例子. 假设 U 是 $\mathbf{F}^3$ 中最后一个坐标为 0 的那些向量所组成的子空间, W 是 $\mathbf{F}^3$ 中前两个坐标为 0 的那些向量所组成的子空间:

$$U = \{(x, y, 0) \in \mathbf{F}^3 : x, y \in \mathbf{F}\}, \quad W = \{(0, 0, z) \in \mathbf{F}^3 : z \in \mathbf{F}\}.$$

那么你应该验证, $\mathbf{F}^3 = U \oplus W$.

又如, 假设 U_j 是 $\mathbf{F}^n$ 中除第 j 个坐标以外其余坐标全是 0 的那些向量所组成的子空间 (例如, $U_2 = \{(0, x, 0, \cdots, 0) \in \mathbf{F}^n : x \in \mathbf{F}\}$), 那么可以证明

$$\mathbf{F}^n = U_1 \oplus \cdots \oplus U_n.$$

现在看最后一个例子. 考虑系数在 $\mathbf{F}$ 中的所有多项式所构成的向量空间 $\mathcal{P}(\mathbf{F})$. 设 U_e 表示由如下形式的多项式 $\boldsymbol{p}$ 所组成的 $\mathcal{P}(\mathbf{F})$ 的子空间

$$\boldsymbol{p}(z) = a_0 + a_2 z^2 + \cdots + a_{2m} z^{2m},$$

并设 U_o 表示由如下形式的多项式 $\boldsymbol{p}$ 所组成的 $\mathcal{P}(\mathbf{F})$ 的子空间

$$\boldsymbol{p}(z) = a_1 z + a_3 z^3 + \cdots + a_{2m+1} z^{2m+1};$$

这里 m 是非负整数, $a_0, \cdots, a_{2m+1} \in \mathbf{F}$ (记号 U_e 和 U_o 可以提醒你 z 的偶次幂和奇次幂). 你应该验证

$$\mathcal{P}(\mathbf{F}) = U_e \oplus U_o.$$

反例有时像例子一样能增加我们的理解. 考虑 $\mathbf{F}^3$ 的如下三个子空间:

$$U_1 = \{(x, y, 0) \in \mathbf{F}^3 : x, y \in \mathbf{F}\};$$

$$U_2 = \{(0, 0, z) \in \mathbf{F}^3 : z \in \mathbf{F}\};$$

$$U_3 = \{(0, y, y) \in \mathbf{F}^3 : y \in \mathbf{F}\}.$$

显然 $\mathbf{F}^3 = U_1 + U_2 + U_3$, 因为任意向量 $(x, y, z) \in \mathbf{F}^3$ 都可以写成

$$(x, y, z) = (x, y, 0) + (0, 0, z) + (0, 0, 0),$$

其中右端的第一个向量含于 U_1, 第二个含于 U_2, 第三个含于 U_3. 然而, $\mathbf{F}^3$ 并非 U_1, U_2, U_3 的直和, 因为向量 $(0, 0, 0)$ 能用两种不同的方式写成和 $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3$, 其中 $\boldsymbol{u}_j \in U_j$. 具体地, 我们有

$$(0, 0, 0) = (0, 1, 0) + (0, 0, 1) + (0, -1, -1),$$

当然还有,

$$(0, 0, 0) = (0, 0, 0) + (0, 0, 0) + (0, 0, 0),$$

其中每个等式右端的第一个向量都含于 U_1, 第二个都含于 U_2, 第三个都含于 U_3.

在上面的例子中, 我们通过证明 $\mathbf{0}$ 表示成适当向量的和时表示法不唯一, 来证明某个和不是直和. 直和的定义要求向量空间中的每个向量都能唯一地表示成一个适当的和. 假设一组子空间的和等于整个空间. 下一个命题表明, 在确定这组子空间是否可以做直和时, 只需考虑 $\mathbf{0}$ 是否可以唯一地写成一个适当的和.

1.8 命题: 设 $U_1, \cdots, U_n$ 都是 V 的子空间, 则 $V = U_1 \oplus \cdots \oplus U_n$ 当且仅当下列两个条件成立:

(a) $V = U_1 + \cdots + U_n$;

(b) 若 $\mathbf{0} = \boldsymbol{u}_1 + \cdots + \boldsymbol{u}_n$, $\boldsymbol{u}_j \in U_j$, 则每个 $\boldsymbol{u}_j$ 都为 $\mathbf{0}$.

证明: 首先假设 $V = U_1 \oplus \cdots \oplus U_n$. 显然 (a) 成立 (根据和与直和的定义). 为证 (b), 假设 $\boldsymbol{u}_1 \in U_1, \cdots, \boldsymbol{u}_n \in U_n$, 并且

$$\mathbf{0} = \boldsymbol{u}_1 + \cdots + \boldsymbol{u}_n,$$

那么每个 $\boldsymbol{u}_j$ 都必为 $\mathbf{0}$ (这是根据直和定义中的唯一性, 以及 $\mathbf{0} = \mathbf{0} + \cdots + \mathbf{0}$, $\mathbf{0} \in U_1, \cdots, \mathbf{0} \in U_n$), 这就证明了 (b).

现在假设 (a) 和 (b) 都成立. 设 $\boldsymbol{v} \in V$, 则由 (a) 可知, 存在 $\boldsymbol{u}_1 \in U_1, \cdots, \boldsymbol{u}_n \in U_n$ 使得

$$\boldsymbol{v} = \boldsymbol{u}_1 + \cdots + \boldsymbol{u}_n.$$

为了证明这个表示是唯一的, 假设还有

$$\boldsymbol{v} = \boldsymbol{v}_1 + \cdots + \boldsymbol{v}_n,$$

其中 $\boldsymbol{v}_1 \in U_1, \cdots, \boldsymbol{v}_n \in U_n$. 把上面的两个等式相减可得

$$\mathbf{0} = (\boldsymbol{u}_1 - \boldsymbol{v}_1) + \cdots + (\boldsymbol{u}_n - \boldsymbol{v}_n).$$

显然 $\boldsymbol{u}_1 - \boldsymbol{v}_1 \in U_1, \cdots, \boldsymbol{u}_n - \boldsymbol{v}_n \in U_n$, 由上式及 (b) 可知每个 $\boldsymbol{u}_j - \boldsymbol{v}_j$ 都为 $\mathbf{0}$. 于是, $\boldsymbol{u}_1 = \boldsymbol{v}_1, \cdots, \boldsymbol{u}_n = \boldsymbol{v}_n$. ■

下一个命题给出了验证两个子空间可以做直和的一个简单条件. 应注意此命题只考虑了两个子空间的情形. 在确定两个以上子空间是否可以做直和时, 只验证任意两个子空间的交为 $\mathbf{0}$ 是不够的. 为了看出这一点, 考虑 1.8 前面的那个反例. 由此反例可知 $\mathbf{F}^3 = U_1 + U_2 + U_3$, 但 $\mathbf{F}^3$ 不是 U_1, U_2, U_3 的直和, 然而 (你应该验证) 却有 $U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{\mathbf{0}\}$. 下一个命题给出了两个子空间可以做直和的一个充分必要条件.

子空间的和类似于子集的并. 同样, 子空间的直和类似于子集的不交并. 任意两个子空间都相交, 因为它们都包含 $\mathbf{0}$. 因此, 交为 $\{\mathbf{0}\}$代替了不相交, 至少在两个子空间的情形如此.

1.9 命题: 设 U 和 W 都是 V 的子空间, 则 $V = U \oplus W$ 当且仅当 $V = U + W$, 并且 $U \cap W = \{\mathbf{0}\}$.

证明: 首先假设 $V = U \oplus W$, 则 $V = U + W$ (由直和的定义). 此外, 若 $\boldsymbol{v} \in U \cap W$, 则 $0 = \boldsymbol{v} + (-\boldsymbol{v})$, 其中 $\boldsymbol{v} \in U$, $-\boldsymbol{v} \in W$. 由于 $\mathbf{0}$ 可唯一地表示成 U 中向量与 W 中向量的和, 从而 $\boldsymbol{v} = \mathbf{0}$. 于是 $U \cap W = \{\mathbf{0}\}$, 这就证明了定理的一个方面.

为证定理的另一个方面, 假设 $V=U+W$, 并且 $U\cap W=\{\mathbf{0}\}$. 为证 $V=U\oplus W$, 假设

$$\mathbf{0}=\boldsymbol{u}+\boldsymbol{w},$$

其中 $\boldsymbol{u}\in U$, $\boldsymbol{w}\in W$. 为了完成证明, 只需证明 $\boldsymbol{u}=\boldsymbol{w}=\mathbf{0}$ (由1.8). 由上面的等式可得 $\boldsymbol{u}=-\boldsymbol{w}\in W$. 于是 $\boldsymbol{u}\in U\cap W$, 故 $\boldsymbol{u}=\mathbf{0}$, 由此及上面的等式可得 $\boldsymbol{w}=\mathbf{0}$, 这就完成了证明. ■

习　题

1. 设 a 和 b 是不全为 0 的实数. 求实数 c 和 d, 使得

$$1/(a+bi)=c+di.$$

2. 证明

$$\frac{-1+\sqrt{3}i}{2}$$

是 1 的一个立方根 (即它的立方等于 1).

3. 证明: 对每个 $\boldsymbol{v}\in V$, 都有 $-(-\boldsymbol{v})=\boldsymbol{v}$.

4. 证明: 若 $a\in\mathbf{F}$, $\boldsymbol{v}\in V$, 且 $a\boldsymbol{v}=\mathbf{0}$, 则 $a=0$ 或者 $\boldsymbol{v}=\mathbf{0}$.

5. 判断 $\mathbf{F}^3$ 的下列子集是不是 $\mathbf{F}^3$ 的子空间:

   (a) $\{(x_1,x_2,x_3)\in\mathbf{F}^3: x_1+2x_2+3x_3=0\}$;
   (b) $\{(x_1,x_2,x_3)\in\mathbf{F}^3: x_1+2x_2+3x_3=4\}$;
   (c) $\{(x_1,x_2,x_3)\in\mathbf{F}^3: x_1x_2x_3=0\}$;
   (d) $\{(x_1,x_2,x_3)\in\mathbf{F}^3: x_1=5x_3\}$.

6. 举出 $\mathbf{R}^2$ 的一个非空子集 U 的例子, 使得 U 对加法和取加法逆封闭 (即当 $\boldsymbol{u}\in U$ 时, $-\boldsymbol{u}\in U$), 但 U 不是 $\mathbf{R}^2$ 的子空间.

7. 举出 $\mathbf{R}^2$ 的一个非空子集 U 的例子, 使得 U 对标量乘法封闭, 但 U 不是 $\mathbf{R}^2$ 的子空间.

8. 证明 V 的任意一组子空间的交都是 V 的一个子空间.

9. 证明 V 的两个子空间的并是 V 的一个子空间当且仅当其中的一个子空间包含在另一个子空间中.

10. 设 U 是 V 的一个子空间. 求 $U+U$.

11. V 的子空间的加法运算具有交换性吗? 结合性呢? (也就是说, 如果 U_1,U_2,U_3 都是 V 的子空间, 是否有 $U_1+U_2=U_2+U_1$? 是否有 $(U_1+U_2)+U_3=U_1+(U_2+U_3)$?)

12. V 的子空间的加法运算有单位元吗? 哪个子空间有加法逆?

13. 证明或举反例：如果 U_1, U_2, W 是 V 的子空间, 使得

$$U_1 + W = U_2 + W,$$

那么 $U_1 = U_2$.

14. 设 U 是由所有形如

$$\boldsymbol{p}(z) = az^2 + bz^5, \quad a, b \in \mathbf{F},$$

的多项式组成的 $\mathcal{P}(\mathbf{F})$ 的一个子空间. 求 $\mathcal{P}(\mathbf{F})$ 的一个子空间 W 使得 $\mathcal{P}(\mathbf{F}) = U \oplus W$.

15. 证明或举反例：如果 U_1, U_2, W 是 V 的子空间, 使得

$$V = U_1 \oplus W, \quad V = U_2 \oplus W,$$

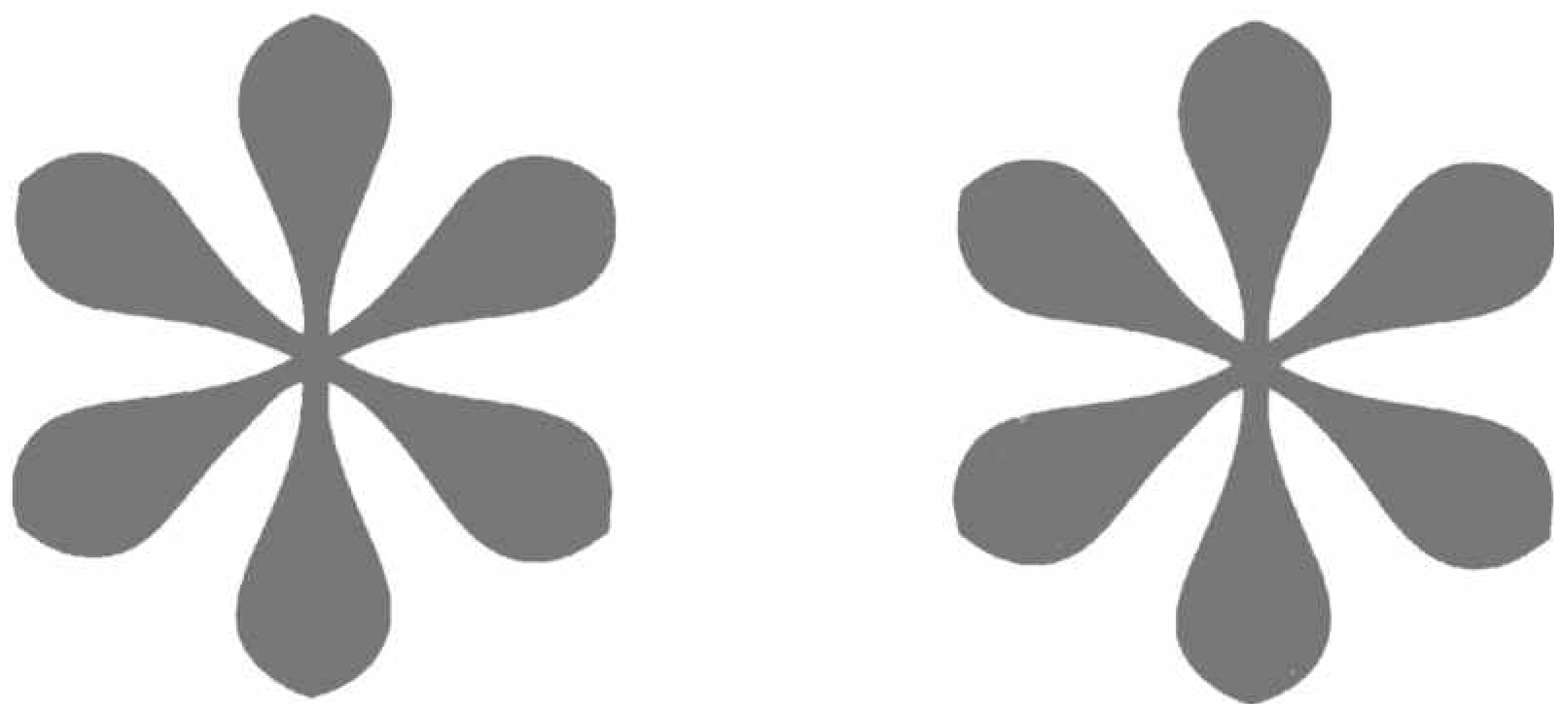
那么 $U_1 = U_2$.

第2章　有限维向量空间

上一章我们学习了向量空间. 线性代数所关注的并不是任意的向量空间, 而是本章所介绍的有限维向量空间. 本章我们将讨论有关这种空间的一些重要概念: 张成、线性无关、基和维数.

回顾一下, 我们总采用如下假定:

$\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$.
V 是 $\mathbf{F}$ 上的向量空间.



§2.1 张成与线性无关

有些数学家采用术语线性张成 (linear span), 意思与张成相同.

V 中一组向量 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 的**线性组合 (linear combination)** 是如下形式的向量

2.1 $$a_1\boldsymbol{v}_1 + \cdots + a_m\boldsymbol{v}_m,$$

其中 $a_1, \cdots, a_m \in \mathbf{F}$. $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 的所有线性组合所构成的集合称为 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 的**张成 (span)**, 记为 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$. 也就是说,

$$\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m) = \{a_1\boldsymbol{v}_1 + \cdots + a_m\boldsymbol{v}_m : a_1, \cdots, a_m \in \mathbf{F}\}.$$

来看关于这些概念的一个例子. 设 $V = \mathbf{F}^3$, 则向量 $(7,2,9)$ 是 $\big((2,1,3), (1,0,1)\big)$ 的一个线性组合, 这是因为

$$(7,2,9) = 2(2,1,3) + 3(1,0,1).$$

于是 $(7,2,9) \in \operatorname{span}\big((2,1,3), (1,0,1)\big)$.

你应该验证, V 中任意一组向量的张成都是 V 的子空间. 为了一致性, 我们声明: 空组 () 的张成等于 $\{\mathbf{0}\}$ (回想一下, 空集不是 V 的子空间).

如果 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 是 V 中一组向量, 那么每个 $\boldsymbol{v}_j$ 都是 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 的线性组合 (为了证明这一点, 令 $a_j = 1$, 并设 2.1 中其他 a 都等于 0). 于是 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 包含每一个 $\boldsymbol{v}_j$. 反之, 由于子空间对标量乘法和加法都封闭, 从而 V 的包含所有 $\boldsymbol{v}_j$ 的子空间必包含 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$. 因此, V 中一组向量的张成是包含这组向量的最小子空间.

如果 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 等于 V, 则称 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ **张成 (span)** V. 如果一个向量空间可以由它的一组向量张成, 则称其为**有限维的 (finite dimensional)**. 例如, 你应该验证, $\mathbf{F}^n$ 是有限维的, 这是因为,

回想一下, 根据定义, 每个组都具有有限长度.

$$\big((1,0,\cdots,0), (0,1,0,\cdots,0), \cdots, (0,\cdots,0,1)\big)$$

张成 $\mathbf{F}^n$.

在给出另一个有限维向量空间的例子之前, 我们需要定义多项式的次数. 对于多项式 $p \in \mathcal{P}(\mathbf{F})$, 如果存在标量 $a_0, a_1, \cdots, a_m \in$

$\mathbf{F}$, $a_m \neq 0$, 使得

2.2 $$p(z) = a_0 + a_1 z + \cdots + a_m z^m, \quad z \in \mathbf{F},$$

则说 p 的**次数 (degree)** 为 m. 规定恒等于 0 的多项式的次数为 $-\infty$.

对于非负整数 m, 令 $\mathcal{P}_m(\mathbf{F})$表示系数在 $\mathbf{F}$ 中并且次数不超过 m 的所有多项式所组成的集合. 你应该验证, $\mathcal{P}_m(\mathbf{F})$ 是 $\mathcal{P}(\mathbf{F})$ 的子空间; 因此 $\mathcal{P}_m(\mathbf{F})$ 是一个向量空间. 这个向量空间是有限维的, 因为它由向量组 $(1, z, \cdots, z^m)$ 张成; 此处我们用 z^k 表示函数 (故 z 是一个哑变量), 这有点滥用记号.

一个向量空间如果不是有限维的, 则称之为**无限维的 (infinite dimensional)**. 例如, $\mathcal{P}(\mathbf{F})$ 是无限维的. 为了证明这个结论, 考虑 $\mathcal{P}(\mathbf{F})$ 中任意一组元素. 记 m 为这组多项式的最高次数 (回想一下, 根据定义, 组的长度是有限的). 则这个组的张成中的每个多项式的次数最多为 m. 因此, 我们所讨论的组不能张成 $\mathcal{P}(\mathbf{F})$. 因为没有组能够张成 $\mathcal{P}(\mathbf{F})$, 所以这个向量空间是无限维的.

我们将不再过多地谈论无限维向量空间, 无限维向量空间是叫作**泛函分析 (functional analysis)** 的数学分支所关注的核心内容. 泛函分析同时采用解析和代数工具.

由 $\mathbf{F}$ 中元素的所有序列组成的向量空间 $\mathbf{F}^\infty$, 也是无限维的, 不过证明有点难. 利用我们即将发展的工具, 你应该可以给出这个证明.

设 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m \in V$, 并且 $\boldsymbol{v} \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$. 由张成的定义, 存在 $a_1, \cdots, a_m \in \mathbf{F}$ 使得

$$\boldsymbol{v} = a_1 \boldsymbol{v}_1 + \cdots + a_m \boldsymbol{v}_m.$$

考虑上式中这些 a 的唯一性问题. 设另一组标量 $\hat{a}_1, \cdots, \hat{a}_m$ 也使得

$$\boldsymbol{v} = \hat{a}_1 \boldsymbol{v}_1 + \cdots + \hat{a}_m \boldsymbol{v}_m.$$

上面这两个等式相减可得

$$\mathbf{0} = (a_1 - \hat{a}_1)\boldsymbol{v}_1 + \cdots + (a_m - \hat{a}_m)\boldsymbol{v}_m.$$

因此, 我们把 $\mathbf{0}$ 写成了 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m$ 的线性组合. 如果 $\mathbf{0}$ 只能用显然的方法 (每个标量都取零) 写成 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m$ 的线性组合, 则每个 $a_j - \hat{a}_j$ 都等于 0, 即每个 a_j 都等于 $\hat{a}_j$ (因此, a 的取法确实是唯一的). 这种情况很重要, 所以我们给它起一个特殊的名字, 这就是我们现在要定义的线性无关性.

对于 V 中一组向量 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$, 如果使得 $a_1\boldsymbol{v}_1+\cdots+a_m\boldsymbol{v}_m$ 等于 $\mathbf{0}$ 的 $a_1,\cdots,a_m\in\mathbf{F}$ 只有 $a_1=\cdots=a_m=0$, 则称 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 是**线性无关的 (linearly independent)**. 例如, 你应该验证,

$$((1,0,0,0),(0,1,0,0),(0,0,1,0))$$

在 $\mathbf{F}^4$ 中是线性无关的. 上一段的推导表明, $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 是线性无关的当且仅当 $\operatorname{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 中每个向量都可以唯一地表示成 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 的线性组合.

大部分线性代数教材定义线性无关集, 而不是线性无关组. 按照线性无关集的定义, 集合 $\{(0,1),(0,1),(1,0)\}$ 在 $\mathbf{F}^2$ 中是线性无关的, 因为它等于集合 $\{(0,1),(1,0)\}$. 而按照我们的定义, 组 $((0,1),(0,1),(1,0))$ 不是线性无关的 (因为 1 乘以第一个向量, 加上 -1 乘以第二个向量, 再加上 0 乘以第三个向量等于 0). 通过讨论组, 而不是讨论集合, 可以避免通常的处理方法所带来的一些问题.

来看线性无关组的另一个例子. 固定一个非负整数 m, 则 $(1,z,\cdots,z^m)$ 在 $\mathcal{P}(\mathbf{F})$ 中是线性无关的. 为了证明这一点, 设 $a_0,a_1,\cdots,a_m\in\mathbf{F}$ 使得

2.3 $$a_0+a_1z+\cdots+a_mz^m=0,\quad z\in\mathbf{F}.$$

如果系数 $a_0,a_1,\cdots,a_m$ 中至少有一个是非零的, 则 z 最多有 m 个不同的值满足 2.3 (如果你对这一事实不熟悉, 可以先承认它; 我们将在第 4 章给出证明); 这一矛盾说明 2.3 中的每个系数都等于 0, 因此, $(1,z,\cdots,z^m)$ 是线性无关的.

V 中的一组向量如果不是线性无关的, 则称为**线性相关的 (linearly dependent)**. 也就是说, V 中一组向量 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 是线性相关的, 当且仅当存在不全为零的 $a_1,\cdots,a_m\in\mathbf{F}$, 使得 $a_1\boldsymbol{v}_1+\cdots+a_m\boldsymbol{v}_m=\mathbf{0}$. 例如, $((2,3,1),(1,-1,2),(7,3,8))$ 在 $\mathbf{F}^3$ 中是线性相关的, 这是因为

$$2(2,3,1)+3(1,-1,2)+(-1)(7,3,8)=(0,0,0)$$

又如, 任意包含 $\mathbf{0}$ 向量的向量组都是线性相关的 (为什么?).

你应该验证, 长度为 1 的组 $(\boldsymbol{v})$ 是线性无关的当且仅当 $\boldsymbol{v}\neq\mathbf{0}$. 你还应该验证, 长度为 2 的组是线性相关的当且仅当其中一个向量是另一个的标量倍. 注意: 上一段中的例子表明, 对于长度为 3 或更大的向量组, 即使其中每一个向量都不是任何其他向量的标量倍, 这个向量组也可能是线性相关的.

你应该验证, 如果从一个线性无关向量组中去掉一些向量, 那么余下的向量组还是线性无关的. 为使这一结论即使在去掉全部向量时仍然成立, 我们声明: 空组 () 是线性无关的.

下面的引理将会经常用到. 它说明, 给定一组线性相关的向量, 其中第一个向量非零, 那么其中必有一个向量包含于它前面

诸向量的张成, 进一步, 我们可以去掉这个向量而不改变原来这组向量的张成.

2.4 线性相关性引理 (Linear Dependent Lemma): 如果 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 在 V 中是线性相关的, 并且 $\boldsymbol{v}_1 \neq \mathbf{0}$, 则有 $j \in \{2, \cdots, m\}$ 使得下列成立:

(a) $\boldsymbol{v}_j \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{j-1})$;

(b) 如果从 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 中去掉第 j 项, 则剩余组的张成等于 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$.

证明: 设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 在 V 中是线性相关的, 并且 $\boldsymbol{v}_1 \neq \mathbf{0}$, 则有不全为 0 的 $a_1, \cdots, a_m \in \mathbf{F}$, 使得

$$a_1\boldsymbol{v}_1 + \cdots + a_m\boldsymbol{v}_m = 0.$$

$a_2, a_3, \cdots, a_m$ 不能全为 0(因为 $\boldsymbol{v}_1 \neq \mathbf{0}$). 设 j 是 $\{2, \cdots, m\}$ 中使得 $a_j \neq 0$ 的最大者, 那么

2.5
$$\boldsymbol{v}_j = -\frac{a_1}{a_j}\boldsymbol{v}_1 - \cdots - \frac{a_{j-1}}{a_j}\boldsymbol{v}_{j-1},$$

这就证明了 (a).

为了证明 (b), 设 $\boldsymbol{u} \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$, 则存在 $c_1, \cdots, c_m \in \mathbf{F}$ 使得

$$\boldsymbol{u} = c_1\boldsymbol{v}_1 + \cdots + c_m\boldsymbol{v}_m.$$

在上面的等式中, 用 2.5 式的右端代替 $\boldsymbol{v}_j$ 可得, $\boldsymbol{u}$ 包含于从 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 中去掉第 j 项所得到的组的张成. 因此 (b) 成立.

■

现在来看一个重要的结果: 线性无关组的长度一定不会大于张成组的长度.

2.6 定理: 在有限维向量空间中, 线性无关向量组的长度小于或等于张成向量组的长度.

如果对每个正整数 m, V 中都存在一个含有 m 个向量的线性无关组, 则该定理表明 V 是无限维的.

证明: 设 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$ 在 V 中是线性无关的, 并且 $(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$ 张成 V. 只需证明 $m \leqslant n$. 我们通过以下步骤来证明. 注意到每一步都添加了一个 $\boldsymbol{u}$, 而去掉了一个 $\boldsymbol{w}$.

第 1 步

由于组 $(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$ 张成 V, 从而再添加任何向量都会得到一个线性相关组. 特别地, 组

$$(\boldsymbol{u}_1, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$$

是线性相关的. 因此利用线性相关引理 (2.4), 我们可以去掉某个 $\boldsymbol{w}$ 而使得由 $\boldsymbol{u}_1$ 和余下的那些 $\boldsymbol{w}$ 构成的组 B (长度为 n) 张成 V.

第 j 步

由于第 $j-1$ 步中的组 B (长度为 n) 张成 V, 从而再添加任何向量都会得到一个线性相关组. 特别地, 在 B 中添加 $\boldsymbol{u}_j$ 于 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{j-1}$ 之后, 那么所得到的长度为 $(n+1)$ 的组是线性相关的. 利用线性相关引理 (2.4), 该组中有一个向量包含于它前面向量的张成, 又因为 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j)$ 是线性无关的, 所以这个向量一定是某个 $\boldsymbol{w}$, 而不是某个 $\boldsymbol{u}$. 我们可以从 B 中去掉这个 $\boldsymbol{w}$, 那么由 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j$ 和余下的那些 $\boldsymbol{w}$ 所构成的新组 B (长度为 n) 张成 V.

经过 m 步, 我们已经添加了所有的 $\boldsymbol{u}$, 程序结束. 如果在其中某一步我们添加了一个 $\boldsymbol{u}$ 但是不再有 $\boldsymbol{w}$ 可以去掉, 就会得到一个矛盾. 因此, 诸 $\boldsymbol{w}$ 至少和诸 $\boldsymbol{u}$ 一样多. ■

直觉告诉我们, 任何包含在有限维向量空间中的向量空间一定也是有限维的. 现在来证明这种直觉是对的.

2.7 命题: *有限维向量空间的子空间都是有限维的.*

证明: 设 V 是有限维向量空间, U 是 V 的一个子空间. 只需证明 U 是有限维的. 我们通过下列几步来证明.

第 1 步

若 $U = \{\mathbf{0}\}$, 则 U 是有限维的, 得证. 若 $U \neq \{\mathbf{0}\}$, 则取非零向量 $\boldsymbol{v}_1 \in U$.

第 j 步

若 $U = \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{j-1})$, 则 U 是有限维的. 若 $U \neq$

$\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{j-1})$, 则取一个向量 $\boldsymbol{v}_j \in U$ 使得

$$\boldsymbol{v}_j \notin \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{j-1}).$$

经过每一步, 只要此程序还在继续, 我们都构造了一个向量组, 使得其中每一个向量都不在它前面向量的张成中. 因此, 由线性相关性引理 (2.4), 经过每一步我们都构造了一个线性无关组. 这一线性无关组不能比 V 的任何张成组长 (由 2.6), 因此, 此程序一定会最终停止, 即 U 是有限维的. ■

§2.2 基

若 V 中一个向量组既是线性无关的又张成 V, 则称之为 V 的**基 (basis)**. 例如,

$$((1, 0, \cdots, 0), (0, 1, 0, \cdots, 0), \cdots, (0, \cdots, 0, 1))$$

是 $\mathbf{F}^n$ 的一个基, 称为 $\mathbf{F}^n$ 的**标准基 (standard basis)**. 除了这个标准基之外, $\mathbf{F}^n$ 还有很多其他的基. 例如, $((1, 2), (3, 5))$ 是 $\mathbf{F}^2$ 的一个基. 组 $((1, 2))$ 是线性无关的, 但不是 $\mathbf{F}^2$ 的基, 因为它不能张成 $\mathbf{F}^2$. 组 $((1, 2), (3, 5), (4, 7))$ 张成 $\mathbf{F}^2$, 但不是基, 因为它不是线性无关的. 又如, $(1, z, \cdots, z^m)$ 是 $\mathcal{P}_m(\mathbf{F})$ 的一个基.

下一个命题表明了基的用处.

2.8 命题: *V 中向量组 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基当且仅当每个 $\boldsymbol{v} \in V$ 都能唯一地写成如下形式*

2.9
$$\boldsymbol{v} = a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n,$$

其中 $a_1, \cdots, a_n \in \mathbf{F}$.

证明: 首先设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, $\boldsymbol{v} \in V$. 因为 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 张成 V, 所以存在 $a_1, \cdots, a_n \in \mathbf{F}$ 使得 2.9 成立. 为了证明 2.9 中的表示是唯一的, 设标量 $b_1, \cdots, b_n$ 使得

$$\boldsymbol{v} = b_1\boldsymbol{v}_1 + \cdots + b_n\boldsymbol{v}_n.$$

这个证明本质上重复了我们定义线性无关性时所采用的思想.

用 2.9 减上面的等式可得

$$\mathbf{0} = (a_1 - b_1)\boldsymbol{v}_1 + \cdots + (a_n - b_n)\boldsymbol{v}_n.$$

这说明每个 $a_j - b_j = 0$ (因为 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是线性无关的), 因此 $a_1 = b_1, \cdots, a_n = b_n$. 这就证明了唯一性, 完成了证明的一个方面.

要证明另一方面, 设每个 $\boldsymbol{v} \in V$ 都可以唯一地写成 2.9 的形式. 显然, 这说明 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 张成 V. 要证明 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是线性无关的, 设 $a_1, \cdots, a_n \in \mathbf{F}$ 使得

$$\mathbf{0} = a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n.$$

由 2.9 中 (对 $\boldsymbol{v} = \mathbf{0}$) 表示的唯一性可得 $a_1 = \cdots = a_n = 0$. 因此 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是线性无关的, 因而是 V 的一个基. ■

向量空间的张成组可能不是基, 因为它可能不是线性无关的. 下一个结果表明: 任给一个张成组, 可以去掉其中的一些向量使得剩余组是线性无关的并且仍然可以张成这个向量空间.

2.10 定理: *在向量空间中, 每个张成组都可以化简成一个基.*

证明: 设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 张成 V, 我们要从 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 中去掉一些向量使得其余向量构成 V 的基. 我们通过以下步骤来完成证明. 从 $B = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 开始.

第 1 步

若 $\boldsymbol{v}_1 = \mathbf{0}$, 则从 B 中去掉 $\boldsymbol{v}_1$. 若 $\boldsymbol{v}_1 \neq \mathbf{0}$, 则保持 B 不变.

第 j 步

若 $\boldsymbol{v}_j$ 属于 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{j-1})$, 则从 B 中去掉 $\boldsymbol{v}_j$. 若 $\boldsymbol{v}_j$ 不属于 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{j-1})$, 则保持 B 不变.

经过 n 步以后终止程序, 得到一个组 B. 因为最初的组张成 V, 而去掉的向量都已经包含于其前面诸向量的张成, 所以这个组 B 张成 V. 这一程序确保 B 中向量都不包含于它前面诸向量的张成. 因此由线性相关性引理 (2.4) 知 B 是线性无关的. 于是 B 是 V 的一个基. ■

考虑组

$$((1,2),(3,6),(4,7),(5,9)),$$

这个组张成 $\mathbf{F}^2$. 为了理解上面的证明, 你应该验证, 如果对这个组应用上面的证明过程, 就得到了 $\mathbf{F}^2$ 的一个基 $((1,2),(4,7))$.

下一个结果是上面定理的一个简单推论, 它说明每个有限维向量空间都有基.

2.11 推论: *每个有限维向量空间都有基.*

证明: 根据定义, 有限维向量空间都有张成组. 前面的定理告诉我们, 任意张成组都可化简成一个基. ■

按照我们的规定, 有限维向量空间 $\{\mathbf{0}\}$ 不是上面推论的反例. 这是因为我们规定空组 () 是线性无关的并且张成 $\{\mathbf{0}\}$, 所以 () 是向量空间 $\{\mathbf{0}\}$ 的一个基.

2.10 说明, 每个张成组都可以化简成基. 在某种意义上说, 下一个定理是 2.10 的一个对偶. 现在我们来证明, 对于任意给定的线性无关组, 都可以添加一些向量使得扩充后的组仍然是线性无关的, 并且还可以张成整个空间.

2.12 定理: *在有限维向量空间中, 每个线性无关向量组都可以扩充成一个基.*

这个定理可以用来给出上面推论的另一个证明. 具体地, 设 V 是有限维的. 这个定理说明空组 () 可以扩充成 V 的一个基. 特别地, V 有基.

证明: 设 V 是有限维的, $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 在 V 中线性无关. 我们想把 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 扩充成 V 的一个基. 我们通过以下步骤来完成证明. 首先设 $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_n)$ 是张成 V 的任意一组向量.

第 1 步

若$\boldsymbol{w}_1$ 含于 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 的张成, 则令 $B=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$. 若 $\boldsymbol{w}_1$ 不含于 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 的张成, 则令 $B=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m,\boldsymbol{w}_1)$.

第 j 步

若$\boldsymbol{w}_j$ 含于 B 的张成, 则保持 B 不变. 若 $\boldsymbol{w}_j$ 不含于 B 的张成, 则通过添加 $\boldsymbol{w}_j$ 来扩充 B.

经过每一步, B 都保持线性无关性, 因为否则的话由线性相关性引理 (2.4) 就会得到矛盾 (回想一下, $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 是线性

无关的, 并且添加到 B 中的任意 $\boldsymbol{w}_j$ 都不包含于 B 中以前的诸向量的张成). 经过 n 步, B 的张成就包含了所有的 $\boldsymbol{w}$. 于是经过 n 步所得到的 B 张成 V, 因此 B 是 V 的一个基. ■

作为上面定理的一个很好的应用, 我们现在来证明, 对于有限维向量空间的每个子空间都可以找到另一个子空间, 使得整个空间是这两个子空间的直和.

即使不假设 V 是有限维的, 利用同样的基本思想, 结合更高级的工具也可以证明这个命题.

2.13 命题: 设 V 是有限维的, U 是 V 的一个子空间, 则存在 V 的一个子空间 W 使得 $V = U \oplus W$.

证明: 因为 V 是有限维的, 所以 U 也是有限维的 (参见 2.7), 从而 U 有一个基 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$ (参见 2.11). $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$ 当然是 V 中的一个线性无关组, 因此可以扩充成 V 的一个基 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$ (参见 2.12). 令 $W = \operatorname{span}(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$.

要证明 $V = U \oplus W$, 只需证明

$$V = U + W, \quad U \cap W = \{\mathbf{0}\};$$

参见 1.9. 要证明第一个等式, 设 $\boldsymbol{v} \in V$. 因为组 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$ 张成 V, 所以存在标量 $a_1, \cdots, a_m, b_1, \cdots, b_n \in \mathbf{F}$ 使得

$$\boldsymbol{v} = \underbrace{a_1\boldsymbol{u}_1 + \cdots + a_m\boldsymbol{u}_m}_{\boldsymbol{u}} + \underbrace{b_1\boldsymbol{w}_1 + \cdots + b_n\boldsymbol{w}_n}_{\boldsymbol{w}}.$$

这就是说 $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}$, 其中 $\boldsymbol{u} \in U$ 和 $\boldsymbol{w} \in W$ 的定义如上. 因此 $\boldsymbol{v} \in U + W$, 这就完成了 $V = U + W$ 的证明.

要证明 $U \cap W = \{\mathbf{0}\}$, 设 $\boldsymbol{v} \in U \cap W$, 则存在标量 $a_1, \cdots, a_m, b_1, \cdots, b_n \in \mathbf{F}$ 使得

$$\boldsymbol{v} = a_1\boldsymbol{u}_1 + \cdots + a_m\boldsymbol{u}_m = b_1\boldsymbol{w}_1 + \cdots + b_n\boldsymbol{w}_n.$$

因此

$$a_1\boldsymbol{u}_1 + \cdots + a_m\boldsymbol{u}_m - b_1\boldsymbol{w}_1 - \cdots - b_n\boldsymbol{w}_n = \mathbf{0}.$$

因为 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$ 是线性无关的, 所以 $a_1 = \cdots = a_m = b_1 = \cdots = b_n = 0$, 从而 $\boldsymbol{v} = \mathbf{0}$, 这就完成了 $U \cap W = \{\mathbf{0}\}$ 的证明. ■

§2.3 维　数

虽然我们已经讨论了有限维向量空间, 但我们还没有定义有限维向量空间的维数. 维数该怎样定义呢? 合理的定义应该使得 $\mathbf{F}^n$ 的维数等于 n. 注意到基

$$((1, 0, \cdots, 0), (0, 1, 0, \cdots, 0), \cdots, (0, \cdots, 0, 1))$$

的长度为 n, 因此我们想把维数定义成基的长度. 但一般来说, 一个给定的有限维向量空间可能有很多不同的基, 而只有当所有基都具有相同长度时, 我们所期望的定义才有意义. 幸好情况就是这样, 我们现在就给出证明.

2.14 定理: *有限维向量空间的任意两个基的长度都相同.*

证明: 设 V 是有限维的, B_1 和 B_2 是 V 的任意两个基, 则 B_1 在 V 中是线性无关的, 并且 B_2 张成 V, 故 B_1 的长度不超过 B_2 的长度 (由 2.6). 互换 B_1 和 B_2 的角色, 可知 B_2 的长度也不超过 B_1 的长度. 因此 B_1 的长度一定等于 B_2 的长度. ■

既然有限维向量空间的任意两个基都具有相同的长度, 我们就可以正式地定义有限维向量空间的维数. 有限维向量空间的任意基的长度称为这个向量空间的**维数 (dimension)**, V 的维数 (如果 V 是有限维的) 记为 $\dim V$. 例如, $\dim \mathbf{F}^n = n$, $\dim \mathcal{P}_m(\mathbf{F}) = m + 1$.

有限维向量空间的每个子空间都是有限维的 (由 2.7), 因此都有维数. 下一个结果所给出的子空间维数的不等式是在预料之中的.

2.15 命题: *若 V 是有限维的, 并且 U 是 V 的子空间, 则 $\dim U \leqslant \dim V$.*

证明: 设 V 是有限维的, 并且 U 是 V 的子空间, 则 U 的任意一个基都是 V 中的一个线性无关向量组, 从而可以扩充成 V 的一个基 (由 2.12). 因此 U 的基的长度小于或等于 V 的基的长度. ■

实向量空间 $\mathbf{R}^2$ 的维数为 2; 复向量空间 $\mathbf{C}$ 的维数为 1. 而作为集合, $\mathbf{R}^2$ 可以和 $\mathbf{C}$ 等同起来 (两个空间上的加法是相同的, 用实数来作标量乘法也一样). 因此, 在讨论向量空间的维数时, $\mathbf{F}$ 所扮演的角色是不能忽略的.

根据定义, 要验证 V 中一组向量是 V 的基, 我们必须证明这个向量组满足两个性质: 它必须是线性无关的, 并且张成 V. 下面两个结果表明, 如果所讨论的组具有适当的长度, 则只需要验证它满足所要求的两个性质之一. 我们先证明每个具有适当长度的张成组都是基.

2.16 命题: *若 V 是有限维的, 则 V 中每个长度为 $\dim V$ 的张成向量组都是 V 的一个基.*

证明: 设 $\dim V = n$, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 张成 V, 则组 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 可以化简成 V 的基 (由 2.10). 然而 V 的每个基的长度都为 n, 所以此处的化简是平凡的, 即 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 中没有任何元素被去掉. 也就是说, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的一个基. ■

现在我们来证明线性无关性足以保证具有适当长度的组都是基.

2.17 命题: *如果 V 是有限维的, 则 V 中每个长度为 $\dim V$ 的线性无关向量组都是 V 的基.*

证明: 设 $\dim V = n$, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 在 V 中是线性无关的, 则组 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 可以扩充成 V 的基 (由 2.12). 然而, V 的每个基的长度都为 n, 所以此处的扩充是平凡的, 即 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 不必添加任何元素. 也就是说, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基. ■

作为上面命题应用的一个例子, 考虑组 $\big((5,7),(4,3)\big)$. $\mathbf{F}^2$ 中这个含有两个向量的组显然是线性无关的, 因为任意向量都不是另外一个向量的标量倍). 因为 $\mathbf{F}^2$ 的维数为 2, 上面的命题表明这个长度为 2 的线性无关组是 $\mathbf{F}^2$ 的一个基 (不必再验证这个组张成 $\mathbf{F}^2$).

下一个定理给出了有限维向量空间的两个子空间之和的维数公式.

2.18 定理: 如果 U_1 和 U_2 是同一个有限维向量空间的两个子空间, 那么

$$\dim(U_1+U_2)=\dim U_1+\dim U_2-\dim(U_1\cap U_2).$$

两个子空间之和的这个维数公式类似于一个熟知的计数公式: 两个有限集合的并集的元素个数等于第一个集合的元素个数, 加上第二个集合的元素个数, 再减去这两个集合的交集的元素个数.

证明: 设 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m)$ 是 $U_1\cap U_2$ 的基, 则 $\dim(U_1\cap U_2)=m$. 因为 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m)$ 是 $U_1\cap U_2$ 的基, 所以它在 U_1 中线性无关, 因此可以扩充成 U_1 的一个基 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j)$ (由 2.12). 于是 $\dim U_1=m+j$. 再将 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m)$ 扩充成 U_2 的一个基 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k)$. 于是 $\dim U_2=m+k$.

现在只需证明 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k)$ 是 U_1+U_2 的一个基, 因为由此可得

$$\begin{aligned}\dim(U_1+U_2)&=m+j+k\\&=(m+j)+(m+k)-m\\&=\dim U_1+\dim U_2-\dim(U_1\cap U_2).\end{aligned}$$

显然 $\operatorname{span}(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k)$ 包含 U_1 和 U_2, 从而包含 U_1+U_2. 因此为了证明这个组是 U_1+U_2 的基, 只需证明它是线性无关的. 为此, 假设

$$a_1\boldsymbol{u}_1+\cdots+a_m\boldsymbol{u}_m+b_1\boldsymbol{v}_1+\cdots+b_j\boldsymbol{v}_j+c_1\boldsymbol{w}_1+\cdots+c_k\boldsymbol{w}_k=\boldsymbol{0},$$

其中所有的 a, b, c 都是标量. 往证所有的标量 a, b, c 都等于 0. 上式可以写成

$$c_1\boldsymbol{w}_1+\cdots+c_k\boldsymbol{w}_k=-a_1\boldsymbol{u}_1-\cdots-a_m\boldsymbol{u}_m-b_1\boldsymbol{v}_1-\cdots-b_j\boldsymbol{v}_j,$$

即 $c_1\boldsymbol{w}_1+\cdots+c_k\boldsymbol{w}_k\in U_1$. 因为所有的 $\boldsymbol{w}$ 都属于 U_2, 所以 $c_1\boldsymbol{w}_1+\cdots+c_k\boldsymbol{w}_k\in U_1\cap U_2$. 又因为 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m)$ 是 $U_1\cap U_2$ 的基, 所以有标量 $d_1,\cdots,d_m$, 使得

$$c_1\boldsymbol{w}_1+\cdots+c_k\boldsymbol{w}_k=d_1\boldsymbol{u}_1+\cdots+d_m\boldsymbol{u}_m.$$

但是 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k)$ 是线性无关的, 故由上式可知, 所有的 c (和 d) 都等于 0. 因此, 最初的那个包含这些 a, b, c 的等式变成

$$a_1\boldsymbol{u}_1+\cdots+a_m\boldsymbol{u}_m+b_1\boldsymbol{v}_1+\cdots+b_j\boldsymbol{v}_j=\mathbf{0}.$$

因为组 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j)$ 是线性无关的, 所以由这个等式可知, 所有的 a, b 都是 0. 于是, 所有的 a, b, c 都等于 0. ■

下一个命题表明维数与直和有很好的关系. 后面的几章会用到这个结果.

回想一下, 直和类似于不交并. 因此 2.19 与下面的陈述是类似的: 如果有限集合 B 可以写成 $A_1\cup\cdots\cup A_m$, 并且这些 A 的元素个数之和等于 B 的元素个数, 那么这个并是不交并.

2.19 命题: 设 V 是有限维的, 并且 $U_1,\cdots,U_m$ 是 V 的子空间, 使得

2.20 $$V=U_1+\cdots+U_m$$

并且

2.21 $$\dim V=\dim U_1+\cdots+\dim U_m.$$

则 $V=U_1\oplus\cdots\oplus U_m$.

证明: 取每个 U_j 的一个基. 将这些基合在一起构成一个长度为 $\dim V$ (由 2.21) 的组, 并且这个组张成 V (由 2.20), 从而是 V 的一个基 (由 2.16). 特别地, 这个组是线性无关的.

现在假设 $\boldsymbol{u}_1\in U_1,\cdots,\boldsymbol{u}_m\in U_m$ 使得

$$\mathbf{0}=\boldsymbol{u}_1+\cdots+\boldsymbol{u}_m.$$

我们可以将每个 $\boldsymbol{u}_j$ 写成 U_j 的 (上面所选取的) 基向量的线性组合. 将这个线性组合带入上面的表达式, 我们就将 $\mathbf{0}$ 写成了 V 的上述基的线性组合. 因此这个线性组合中所用到的标量一定都为 0. 于是每个 $\boldsymbol{u}_j=\mathbf{0}$, 这就证明了 $V=U_1\oplus\cdots\oplus U_m$ (由 1.8). ■

习　题

1. 证明: 如果 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 张成 V, 那么由每个向量 (最后一个向量除外) 减去其后一个向量所得到的组

$$(\boldsymbol{v}_1 - \boldsymbol{v}_2, \boldsymbol{v}_2 - \boldsymbol{v}_3, \cdots, \boldsymbol{v}_{n-1} - \boldsymbol{v}_n, \boldsymbol{v}_n)$$

也张成 V.

2. 证明: 如果 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 在 V 中是线性无关的, 那么由每个向量 (最后一个向量除外) 减去其后一个向量所得到的组

$$(\boldsymbol{v}_1 - \boldsymbol{v}_2, \boldsymbol{v}_2 - \boldsymbol{v}_3, \cdots, \boldsymbol{v}_{n-1} - \boldsymbol{v}_n, \boldsymbol{v}_n)$$

也是线性无关的.

3. 设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 在 V 中是线性无关的, 并且 $\boldsymbol{w} \in V$. 证明: 若 $(\boldsymbol{v}_1 + \boldsymbol{w}, \cdots, \boldsymbol{v}_n + \boldsymbol{w})$ 是线性相关的, 则 $\boldsymbol{w} \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$.

4. 设 m 是正整数. 由 0 和系数在 $\mathbf{F}$ 中次数等于 m 的所有多项式所组成的集合是 $\mathcal{P}(\mathbf{F})$ 的子空间吗?

5. 证明 $\mathbf{F}^\infty$ 是无限维的.

6. 证明由 $[0,1]$ 区间上所有连续的实值函数所组成的实向量空间是无限维的.

7. 证明: V 是无限维的当且仅当 V 中有一个向量序列 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots$, 使得对每个正整数 n, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 都是线性无关的.

8. 令 U 是 $\mathbf{R}^5$ 的由

$$U = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbf{R}^5 : x_1 = 3x_2 \text{ and } x_3 = 7x_4\}$$

定义的子空间. 求 U 的一个基.

9. 证明或反驳: $\mathcal{P}_3(\mathbf{F})$ 有一个基 $(\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3)$ 使得多项式 $\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3$ 的次数都不等于 2.

10. 设 V 是有限维的, $\dim V = n$. 证明在 V 中存在一维子空间 $U_1, \cdots, U_n$ 使得

$$V = U_1 \oplus \cdots \oplus U_n.$$

11. 设 V 是有限维的, U 是 V 的子空间使得 $\dim U = \dim V$. 证明 $U = V$.

12. 设 $\boldsymbol{p}_0, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$ 是 $\mathcal{P}_m(\mathbf{F})$ 中多项式, 使得对任意 j 都有 $\boldsymbol{p}_j(2) = 0$. 证明 $(\boldsymbol{p}_0, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_m)$ 在 $\mathcal{P}_m(\mathbf{F})$ 中不是线性无关的.

13. 设 U 和 W 都是 $\mathbf{R}^8$ 的子空间, 使得 $\dim U = 3$, $\dim W = 5$, 并且 $U + W = \mathbf{R}^8$. 证明 $U \cap W = \{\mathbf{0}\}$.

14. 设 U 和 W 都是 $\mathbf{R}^9$ 的 5 维子空间. 证明 $U \cap W \neq \{\mathbf{0}\}$.

15. 通过与有限集合中三个子集之并的元素个数公式相类比, 你可能会猜到, 如果 U_1, U_2, U_3 是有限维向量空间的子空间, 那么

$$\begin{aligned}\dim(U_1 + U_2 + U_3) = &\dim U_1 + \dim U_2 + \dim U_3 \\ &- \dim(U_1 \cap U_2) - \dim(U_1 \cap U_3) \\ &- \dim(U_2 \cap U_3) \\ &+ \dim(U_1 \cap U_2 \cap U_3).\end{aligned}$$

证明或举反例.

16. 证明: 若 V 是有限维向量空间, 并且 $U_1, \cdots, U_m$ 都是 V 的子空间, 则

$$\dim(U_1 + \cdots + U_m) \leqslant \dim U_1 + \cdots + \dim U_m.$$

17. 设 V 是有限维的. 证明: 如果 $U_1, \cdots, U_m$ 是 V 的子空间, 使得 $V = U_1 \oplus \cdots \oplus U_m$, 那么

$$\dim V = \dim U_1 + \cdots + \dim U_m.$$

这个习题深化了子空间的直和与子集的不交并之间的类比. 特别地, 把这个习题与下面的明显陈述相比较: 如果一个有限集合写成了子集的不交并, 那么该集合的元素个数等于这些不相交子集的元素个数之和.

第3章　线性映射

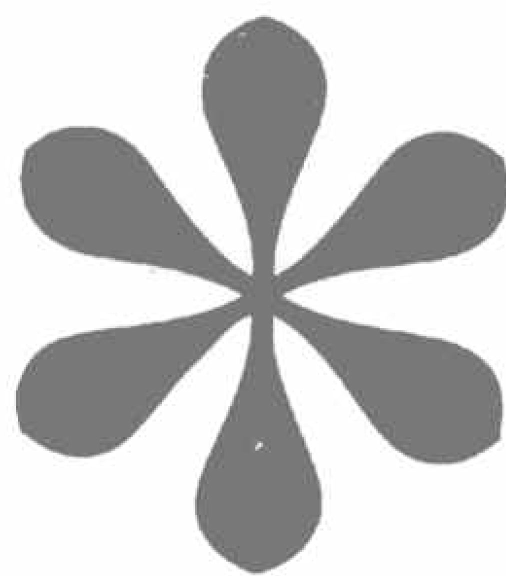
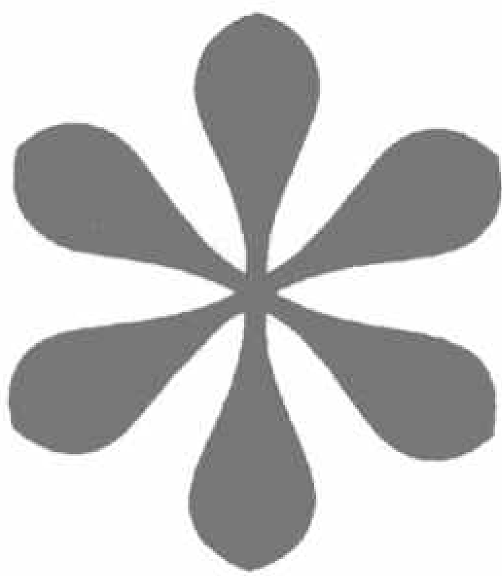
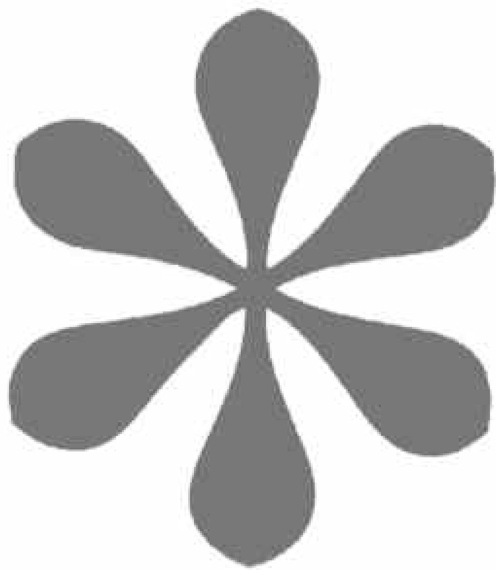
迄今我们所关注的都是向量空间, 而向量空间并不那么让人兴奋. 我们现在要讨论的主题 —— 线性映射才是真正让人感兴趣的部分.

先回顾一下, 我们总采用如下假定:

$\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$.
V 是 $\mathbf{F}$ 上的向量空间.

在这一章, 我们还经常需要 V 之外的另一个向量空间 W:

在本章的其余部分,
W 表示 $\mathbf{F}$ 上的向量空间.

§3.1　定义与例子

有些数学家使用**线性变换** (linear transformation) 这样的术语，意思与线性映射相同.

从 V 到 W 的**线性映射** (linear map) 是具有下列性质的函数 $T : V \to W$:

加性 (addivity)

对所有 $\boldsymbol{u}, \boldsymbol{v} \in V$ 都有 $T(\boldsymbol{u}+\boldsymbol{v}) = T\boldsymbol{u} + T\boldsymbol{v}$;

齐性 (homogeneity)

对所有 $a \in \mathbf{F}$, $\boldsymbol{v} \in V$ 都有 $T(a\boldsymbol{v}) = a(T\boldsymbol{v})$.

注意, 对于线性映射, 我们经常使用记号 $T\boldsymbol{v}$, 也使用更标准的函数记号 $T(\boldsymbol{v})$.

从 V 到 W 的所有线性映射所构成的集合记为 $\mathcal{L}(V, W)$. 我们来看线性映射的一些例子. 你一定要验证下面所定义的函数的确是线性映射.

零 (zero)

除了其他用法之外, 我们还用符号 0 表示一个函数, 它把某个向量空间的每个元素都映成另一个向量空间的加法单位元. 确切地说, $0 \in \mathcal{L}(V, W)$ 定义如下

$$0\boldsymbol{v} = \mathbf{0}.$$

注意, 等式左边的 0 是从 V 到 W 的函数, 而右边的 $\mathbf{0}$ 是 W 的加法单位元. 一般来说, 通过上下文可以辨别符号 0 的各种用法.

恒等 (identity)

恒等映射 (identity map) 是某个向量空间上的函数, 记为 I, 它把每个元素都映成自身. 确切地说, $I \in \mathcal{L}(V, V)$ 定义如下

$$I\boldsymbol{v} = \boldsymbol{v}.$$

微分 (differentiation)

定义 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 如下

$$T\boldsymbol{p} = \boldsymbol{p}'.$$

这个函数是线性的, 此论断只是把关于微分的一个基本结

果换了一种说法: $(f+g)'=f'+g'$, $(af)'=af'$, 其中 f,g 是可微的, a 是常数.

积分 (integration)

定义 $T\in\mathcal{L}(\mathcal{P}(\mathbf{R}),\mathbf{R})$ 如下

$$T\boldsymbol{p}=\int_0^1\boldsymbol{p}(x)\mathrm{d}x.$$

这个函数是线性的, 此论断只是把关于积分的一个基本结果换了一种说法: 两个函数之和的积分等于这两个函数积分的和, 常数与函数乘积的积分等于常数乘以函数的积分.

x^2 乘 (multiplication by x^2)

定义 $T\in\mathcal{L}(\mathcal{P}(\mathbf{R}),\mathcal{P}(\mathbf{R}))$ 如下

$$(T\boldsymbol{p})(x)=x^2\boldsymbol{p}(x),\quad x\in\mathbf{R}.$$

后向移位 (backward shift)

回忆一下, $\mathbf{F}^\infty$ 表示 $\mathbf{F}$ 中元素的所有序列所组成的向量空间. 定义 $T\in\mathcal{L}(\mathbf{F}^\infty,\mathbf{F}^\infty)$ 如下

$$T(x_1,x_2,x_3,\cdots)=(x_2,x_3,\cdots).$$

从 $\mathbf{F}^n$ 到 $\mathbf{F}^m$ (from $\mathbf{F}^n$ to $\mathbf{F}^m$)

定义 $T\in\mathcal{L}(\mathbf{R}^3,\mathbf{R}^2)$ 如下

$$T(x,y,z)=(2x-y+3z,7x+5y-6z).$$

更一般地, 设 m 和 n 都是正整数, $a_{j,k}\in\mathbf{F}$, $j=1,\cdots,m$, $k=1,\cdots,n$. 定义 $T\in\mathcal{L}(\mathbf{F}^n,\mathbf{F}^m)$ 如下

$$\begin{aligned}T(x_1,\cdots,x_n)=&(a_{1,1}x_1+\cdots+a_{1,n}x_n,\cdots,\\&a_{m,1}x_1+\cdots+a_{m,n}x_n).\end{aligned}$$

以后我们会看到从 $\mathbf{F}^n$ 到 $\mathbf{F}^m$ 的每个线性映射都是这种形式的.

尽管线性映射在整个数学中是广泛存在的, 但某些学生却误解为线性映射无处不在, 这些学生似乎认为 cos 是 $\mathbf{R}$ 到 $\mathbf{R}$ 的线性映射而写出了像 $\cos 2x=2\cos x$ 和 $\cos(x+y)=\cos x+\cos y$ 这样的“恒等式”.

设 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的一个基, 并且 $T:V\to W$ 是线性的. 如果 $\boldsymbol{v}\in V$, 那么 $\boldsymbol{v}$ 可写成如下形式

$$\boldsymbol{v}=a_1\boldsymbol{v}_1+\cdots+a_n\boldsymbol{v}_n.$$

由 T 的线性可知

$$Tv = a_1Tv_1 + \cdots + a_nTv_n.$$

特别地, $Tv_1, \cdots, Tv_n$ 的值确定了 T 在 V 的任意向量上的值.

线性映射可根据其在一个基上的取值来构造, 而这些取值可以是任意指定的. 具体地, 给定 V 的基 $(v_1, \cdots, v_n)$ 和任意指定的向量 $w_1, \cdots, w_n \in W$, 我们都可以构造一个线性映射 $T: V \to W$ 使得 $Tv_j = w_j$, $j = 1, \cdots, n$. 这样就别无选择了, 必须定义 T 如下:

$$T(a_1v_1 + \cdots + a_nv_n) = a_1w_1 + \cdots + a_nw_n,$$

其中 $a_1, \cdots, a_n$ 是 $\mathbf{F}$ 中的任意元素. 由于 $(v_1, \cdots, v_n)$ 是 V 的基, 所以上面的等式的确定义了从 V 到 W 的函数 T. 你应该验证, 如上定义的函数 T 是线性的, 并且 $Tv_j = w_j$, $j = 1, \cdots, n$.

现在, 我们要在 $\mathcal{L}(V, W)$ 上定义加法和标量乘法, 以使其成为一个向量空间. 对于 $S, T \in \mathcal{L}(V, W)$, 采用通常的函数加法, 可以定义一个函数 $S + T \in \mathcal{L}(V, W)$:

$$(S + T)v = Sv + Tv, \quad v \in V.$$

你应该验证, 当 $S, T \in \mathcal{L}(V, W)$ 时, $S + T$ 的确是从 V 到 W 的线性映射. 对于 $a \in \mathbf{F}$ 和 $T \in \mathcal{L}(V, W)$, 采用通常的数与函数的乘法, 可以定义一个函数 $aT \in \mathcal{L}(V, W)$:

$$(aT)v = a(Tv), \quad v \in V.$$

你应该验证, 当 $a \in \mathbf{F}$, $T \in \mathcal{L}(V, W)$ 时, aT 的确是从 V 到 W 的线性映射. 对于刚才定义的这些运算, (你应该验证) $\mathcal{L}(V, W)$ 成为一个向量空间. 注意, $\mathcal{L}(V, W)$ 的加法单位元就是本节前面所定义的零映射.

一般来说, 向量空间中的两个元素相乘是没有意义的, 但是对于一对适当的线性映射却存在一种有用的乘积. 我们还需要第三个向量空间, 所以假设 U 是 $\mathbf{F}$ 上的向量空间. 如果 $T \in \mathcal{L}(U, V)$, $S \in \mathcal{L}(V, W)$, 那么定义 $ST \in \mathcal{L}(U, W)$ 如下:

$$(ST)(v) = S(Tv), \quad v \in U.$$

也就是说, ST 恰为通常的函数复合 $S \circ T$. 但是, 若两个函数都是线性的,则大多数数学家都写成 ST 而不是 $S \circ T$. 你应该验

证, 当 $T \in \mathcal{L}(U,V)$, $S \in \mathcal{L}(V,W)$ 时, ST 的确是从 U 到 W 的线性映射. 注意, 只有当 T 映到 S 的定义域内时 ST 才有定义. 我们称 ST 是 S 和 T 的**乘积 (product)**. 你应该验证, 它具有乘积的大多数常见性质:

结合性(associativity)

$(T_1T_2)T_3 = T_1(T_2T_3)$, 这里 T_1, T_2, T_3 都是线性映射, 并且乘积都有意义 (即 T_3 必须映到 T_2 的定义域内, 并且 T_2 必须映到 T_1 的定义域内).

恒等映射(identity)

$TI = T$, $IT = T$, 这里 $T \in \mathcal{L}(V,W)$ (注意, 在第一个等式中, I 是 V 上的恒等映射, 而在第二个等式中 I 是 W 上的恒等映射).

分配性质(distributive properties)

$(S_1 + S_2)T = S_1T + S_2T$, $\quad S(T_1 + T_2) = ST_1 + ST_2$, 这里 $T, T_1, T_2 \in \mathcal{L}(U,V)$, $S, S_1, S_2 \in \mathcal{L}(V,W)$.

线性映射的乘法不是交换的. 也就是说, $ST = TS$ 未必成立, 即便是这个等式的两边都有意义. 例如, 若 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 是本节前面所定义的微分映射, $S \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 是本节前面所定义的 x^2 乘映射, 则

$$((ST)\boldsymbol{p})(x) = x^2\boldsymbol{p}'(x), \text{ 但 } ((TS)\boldsymbol{p})(x) = x^2\boldsymbol{p}'(x) + 2x\boldsymbol{p}(x).$$

也就是说, 先乘 x^2 再微分和先微分再乘 x^2 是不同的.

§3.2 零空间与值域

对于 $T \in \mathcal{L}(V,W)$, V 中被 T 映成 $\mathbf{0}$ 的那些向量所组成的子集称为 T 的**零空间 (null space)**, 记为 $\operatorname{null} T$:

$$\operatorname{null} T = \{\boldsymbol{v} \in V : T\boldsymbol{v} = \mathbf{0}\}.$$

有些数学家使用术语**核 (kernel)** 而不是零空间.

来看前一节的几个例子. 在微分的例子中, 我们用 $T\boldsymbol{p} = \boldsymbol{p}'$ 定义了 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$. 只有常函数的导数才是零函数, 于是, 在此情况下, T 的零空间等于常函数之集.

在 x^2 乘的例子中, 我们用 $(T\boldsymbol{p})(x) = x^2\boldsymbol{p}(x)$ 定义了 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$. 满足 $x^2\boldsymbol{p}(x) = \mathbf{0}(x \in \mathbf{R})$ 的多项式 $\boldsymbol{p}$ 只有 $\mathbf{0}$ 多项式. 于是, 在这种情况下我们有

$$\operatorname{null} T = \{\mathbf{0}\}.$$

在后向移位的例子中, 我们用

$$T(x_1, x_2, x_3, \cdots) = (x_2, x_3, \cdots)$$

定义了 $T \in \mathcal{L}(\mathbf{F}^\infty, \mathbf{F}^\infty)$. 显然, $T(x_1, x_2, x_3, \cdots)$ 等于 $\mathbf{0}$ 当且仅当 $x_2, x_3, \cdots$ 都是 0. 于是, 在这种情况下我们有

$$\operatorname{null} T = \{(a, 0, 0, \cdots) : a \in \mathbf{F}\}.$$

下一个命题证明了线性映射的零空间是定义域的子空间. 特别地, $\mathbf{0}$ 包含于每个线性映射的零空间.

3.1 命题: *若 $T \in \mathcal{L}(V, W)$, 则 $\operatorname{null} T$ 是 V 的子空间.*

证明: 设 $T \in \mathcal{L}(V, W)$. 由加性可得

$$T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0}),$$

从而 $T(\mathbf{0}) = \mathbf{0}$. 于是 $\mathbf{0} \in \operatorname{null} T$.

如果 $\boldsymbol{u}, \boldsymbol{v} \in \operatorname{null} T$, 那么

$$T(\boldsymbol{u} + \boldsymbol{v}) = T\boldsymbol{u} + T\boldsymbol{v} = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

从而 $\boldsymbol{u} + \boldsymbol{v} \in \operatorname{null} T$. 于是 $\operatorname{null} T$ 对加法封闭.

如果 $\boldsymbol{u} \in \operatorname{null} T$, $a \in \mathbf{F}$, 那么

$$T(a\boldsymbol{u}) = aT\boldsymbol{u} = a\mathbf{0} = \mathbf{0},$$

从而 $a\boldsymbol{u} \in \operatorname{null} T$. 于是 $\operatorname{null} T$ 对标量乘法封闭.

我们已经证明了 T 的零空间包含 $\mathbf{0}$, 并且对加法和标量乘法都封闭. 因此, $\operatorname{null} T$ 是 V 的子空间. ■

线性映射 $T: V \to W$ 称为**单的** (injective), 如果当 $\boldsymbol{u}, \boldsymbol{v} \in V$, $T\boldsymbol{u} = T\boldsymbol{v}$ 时, 必有 $\boldsymbol{u} = \boldsymbol{v}$. 下一个命题表明, 为了验证线性映射是单的, 只需验证 $\mathbf{0}$ 是唯一一个被映成 $\mathbf{0}$ 的向量. 作为这个命题的一个简单应用, 我们看到, 本节前面计算零空间的三个线性映射 (微分、x^2 乘、后向移位) 中只有 x^2 乘是单的.

很多数学家采用**一对一的** (one-to-one) 这样的术语, 意思与单的相同.

3.2 命题: 设 $T \in \mathcal{L}(V, W)$, 则 T 是单的当且仅当 $\operatorname{null} T = \{\mathbf{0}\}$.

证明: 首先假设 T 是单的. 我们要证明 $\operatorname{null} T = \{\mathbf{0}\}$. 已知 $\{\mathbf{0}\} \subset \operatorname{null} T$ (由 3.1). 为了证明另一个方向的包含关系, 设 $\boldsymbol{v} \in \operatorname{null} T$, 则

$$T(\boldsymbol{v}) = \mathbf{0} = T(\mathbf{0}).$$

因为 T 是单的, 故由上式可得 $\boldsymbol{v} = \mathbf{0}$. 于是 $\operatorname{null} T = \{\mathbf{0}\}$.

反之, 假设 $\operatorname{null} T = \{\mathbf{0}\}$. 我们要证明 T 是单的. 为此, 设 $\boldsymbol{u}, \boldsymbol{v} \in V$, 并且 $T\boldsymbol{u} = T\boldsymbol{v}$, 那么

$$\mathbf{0} = T\boldsymbol{u} - T\boldsymbol{v} = T(\boldsymbol{u} - \boldsymbol{v}).$$

于是 $\boldsymbol{u} - \boldsymbol{v}$ 含于 $\operatorname{null} T = \{\mathbf{0}\}$. 因此 $\boldsymbol{u} - \boldsymbol{v} = \mathbf{0}$, 即 $\boldsymbol{u} = \boldsymbol{v}$. 故 T 是单的. ■

对于 $T \in \mathcal{L}(V, W)$, 由 W 中形如 $T\boldsymbol{v}(\boldsymbol{v} \in V)$ 的向量所组成的子集称为 T 的**值域** (range), 记为 $\operatorname{range} T$:

某些数学家使用**象** (image) 这个词, 意思与值域相同.

$$\operatorname{range} T = \{T\boldsymbol{v} : \boldsymbol{v} \in V\}.$$

例如, 若 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 是由 $T\boldsymbol{p} = \boldsymbol{p}'$ 所定义的微分映射, 则 $\operatorname{range} T = \mathcal{P}(\mathbf{R})$. 这是因为, 对于每个多项式 $\boldsymbol{q} \in \mathcal{P}(\mathbf{R})$ 都存在多项式 $\boldsymbol{p} \in \mathcal{P}(\mathbf{R})$ 使得 $\boldsymbol{p}' = \boldsymbol{q}$.

又如, 若 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 是由 $(T\boldsymbol{p})(x) = x^2\boldsymbol{p}(x)$ 定义的 x^2 乘线性映射, 则 T 的值域是形如 $a_2x^2 + \cdots + a_mx^m$ 的多项式所组成的集合, 其中 $a_2, \cdots, a_m \in \mathbf{R}$.

下一个命题证明了线性映射的值域是目标空间的子空间.

3.3 命题: 如果 $T \in \mathcal{L}(V, W)$, 那么 $\operatorname{range} T$ 是 W 的子空间.

证明: 设 $T \in \mathcal{L}(V, W)$, 则 $T(\mathbf{0}) = \mathbf{0}$ (由 3.1), 即 $\mathbf{0} \in \operatorname{range} T$.

如果 $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \operatorname{range} T$, 那么存在 $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ 使得 $T\boldsymbol{v}_1 = \boldsymbol{w}_1$, $T\boldsymbol{v}_2 = \boldsymbol{w}_2$. 于是

$$T(\boldsymbol{v}_1 + \boldsymbol{v}_2) = T\boldsymbol{v}_1 + T\boldsymbol{v}_2 = \boldsymbol{w}_1 + \boldsymbol{w}_2,$$

故 $\boldsymbol{w}_1 + \boldsymbol{w}_2 \in \operatorname{range} T$. 因此, $\operatorname{range} T$ 对加法封闭.

如果 $\boldsymbol{w} \in \operatorname{range} T$, $a \in \mathbf{F}$, 那么存在 $\boldsymbol{v} \in V$ 使得 $T\boldsymbol{v} = \boldsymbol{w}$. 于是

$$T(a\boldsymbol{v}) = aT\boldsymbol{v} = a\boldsymbol{w},$$

故 $a\boldsymbol{w} \in \operatorname{range} T$. 因此, $\operatorname{range} T$ 对标量乘法封闭.

我们已经证明了 $\operatorname{range} T$ 包含 $\mathbf{0}$, 并且对加法和标量乘法都封闭, 所以 $\operatorname{range} T$ 是 W 的子空间. ■

许多数学家采用**映上 (onto)** 这个术语, 意思与满的相同.

线性映射 $T : V \to W$ 称为**满的 (surjective)**, 如果它的值域等于 W. 例如, 由 $T\boldsymbol{p} = \boldsymbol{p}'$ 定义的微分映射 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 是满的, 因为它的值域等于 $\mathcal{P}(\mathbf{R})$. 又如, 由 $(T\boldsymbol{p})(x) = x^2\boldsymbol{p}(x)$ 定义的线性映射 $T \in \mathcal{L}(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R}))$ 不是满的, 因为它的值域不等于 $\mathcal{P}(\mathbf{R})$. 再如, 你应该验证, 如下定义的后向移位映射 $T \in \mathcal{L}(\mathbf{F}^\infty, \mathbf{F}^\infty)$ 是满的,

$$T(x_1, x_2, x_3, \cdots) = (x_2, x_3, \cdots).$$

线性映射是不是满的与我们把什么看作目标空间有关. 例如, 固定一个正整数 m, 由 $T\boldsymbol{p}=\boldsymbol{p}'$ 定义的微分映射 $T\in\mathcal{L}(\mathcal{P}_m(\mathbf{R}), \mathcal{P}_m(\mathbf{R}))$ 不是满的, 因为多项式 x^m 不包含于 T 的值域. 然而, 由 $T\boldsymbol{p} = \boldsymbol{p}'$ 定义的微分映射 $T \in \mathcal{L}(\mathcal{P}_m(\mathbf{R}), \mathcal{P}_{m-1}(\mathbf{R}))$ 是满的, 因为它的值域等于目标空间 $\mathcal{P}_{m-1}(\mathbf{R})$.

下一个定理是本章的主要结果: 有限维向量空间上的线性映射的零空间的维数加上值域的维数等于定义域的维数.

3.4 定理: *如果 V 是有限维向量空间, 并且 $T \in \mathcal{L}(V, W)$, 那么 $\operatorname{range} T$ 是 W 的有限维子空间, 并且*

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

证明: 假设 V 是有限维向量空间, $T\in\mathcal{L}(V,W)$, $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m)$ 是 $\operatorname{null}T$ 的一个基, 那么 $\dim\operatorname{null}T=m$. 线性无关组 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m)$ 可以扩充成 V 的基 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_n)$ (由 2.12). 于是 $\dim V=m+n$. 为了完成证明, 只需证明 $\operatorname{range}T$ 是有限维的, 并且 $\dim\operatorname{range}T=n$. 为此, 往证 $(T\boldsymbol{w}_1,\cdots,T\boldsymbol{w}_n)$ 是 $\operatorname{range}T$ 的基.

设 $\boldsymbol{v}\in V$. 因为 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_n)$ 张成 V, 所以

$$\boldsymbol{v}=a_1\boldsymbol{u}_1+\cdots+a_m\boldsymbol{u}_m+b_1\boldsymbol{w}_1+\cdots+b_n\boldsymbol{w}_n,$$

其中的这些 a 和 b 都含于 $\mathbf{F}$. 用 T 作用于上式两端可得

$$T\boldsymbol{v}=b_1T\boldsymbol{w}_1+\cdots+b_nT\boldsymbol{w}_n,$$

其中没有出现形如 $T\boldsymbol{u}_j$ 的项, 这是因为每个 $\boldsymbol{u}_j\in\operatorname{null}T$. 上面这个等式表明 $(T\boldsymbol{w}_1,\cdots,T\boldsymbol{w}_n)$ 张成 $\operatorname{range}T$. 特别地, $\operatorname{range}T$ 是有限维的.

为了证明 $(T\boldsymbol{w}_1,\cdots,T\boldsymbol{w}_n)$ 是线性无关的, 设 $c_1,\cdots,c_n\in\mathbf{F}$, 并且

$$c_1T\boldsymbol{w}_1+\cdots+c_nT\boldsymbol{w}_n=\mathbf{0},$$

那么

$$T(c_1\boldsymbol{w}_1+\cdots+c_n\boldsymbol{w}_n)=\mathbf{0},$$

故

$$c_1\boldsymbol{w}_1+\cdots+c_n\boldsymbol{w}_n\in\operatorname{null}T.$$

因为 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m)$ 张成 $\operatorname{null}T$, 所以

$$c_1\boldsymbol{w}_1+\cdots+c_n\boldsymbol{w}_n=d_1\boldsymbol{u}_1+\cdots+d_m\boldsymbol{u}_m,$$

其中的这些 d 都含于 $\mathbf{F}$. 这个等式表明, 所有的 c (和 d) 都是 0 (因为 $(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m,\boldsymbol{w}_1,\cdots,\boldsymbol{w}_n)$ 是线性无关的). 于是, $(T\boldsymbol{w}_1,\cdots,T\boldsymbol{w}_n)$ 是线性无关的, 于是它是 $\operatorname{range}T$ 的基. ■

现在我们可以证明, 从一个有限维向量空间到比它更小的向量空间的线性映射不可能是单的, 这里“更小”是用维数来衡量的.

3.5 推论: 如果 V 和 W 都是有限维向量空间, 并且 $\dim V > \dim W$, 那么 V 到 W 的线性映射一定不是单的.

证明: 假设 V 和 W 都是有限维向量空间, 并且 $\dim V > \dim W$. 对于 $T \in \mathcal{L}(V, W)$, 由 3.4 可得

$$\begin{aligned}\dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T \\ &\geqslant \dim V - \dim W \\ &> 0,\end{aligned}$$

这就证明了 $\dim \operatorname{null} T > 0$, 即 $\operatorname{null} T$ 必须包含非零向量. 因此 T 不是单的 (由 3.2). ■

下一个推论在某种意义上是推论 3.5 的对偶, 它表明从一个有限维向量空间到比它更大的向量空间的线性映射不可能是满的, 这里“更大”是用维数来衡量的.

3.6 推论: 如果 V 和 W 都是有限维向量空间, 并且 $\dim V < \dim W$, 那么 V 到 W 的线性映射一定不是满的.

证明: 假设 V 和 W 都是有限维向量空间, 并且 $\dim V < \dim W$. 对于 $T \in \mathcal{L}(V, W)$, 由 3.4 可得

$$\begin{aligned}\dim \operatorname{range} T &= \dim V - \dim \operatorname{null} T \\ &\leqslant \dim V \\ &< \dim W,\end{aligned}$$

这就证明了 $\dim \operatorname{range} T < \dim W$, 即 $\operatorname{range} T$ 不等于 W. 因此 T 不是满的. ■

上面这两个推论在线性方程组理论中有一些重要结论. 为了看出这一点, 固定正整数 m 和 n, 并设 $a_{j,k} \in \mathbf{F}$, $j = 1, \cdots, m$; $k = 1, \cdots, n$. 定义 $T : \mathbf{F}^n \to \mathbf{F}^m$ 如下

$$T(x_1,\cdots,x_n)=\left(\sum_{k=1}^{n}a_{1,k}x_k,\cdots,\sum_{k=1}^{n}a_{m,k}x_k\right).$$

现在考虑方程 $T\boldsymbol{x}=\boldsymbol{0}$ (其中 $\boldsymbol{x}\in\mathbf{F}^n$, $\boldsymbol{0}$ 是 $\mathbf{F}^m$ 的加法单位元, 即, 由 0 组成的长度为 m 的组). 设 $\boldsymbol{x}=(x_1,\cdots,x_n)$, 那么方程 $T\boldsymbol{x}=\boldsymbol{0}$ 可以重新写成一个齐次方程组:

这里齐次 (homogeneous) 的意思是每个方程右边的常数项都等于 0.

$$\begin{gathered}\sum_{k=1}^{n}a_{1,k}x_k=0\\ \vdots\\ \sum_{k=1}^{n}a_{m,k}x_k=0.\end{gathered}$$

把其中的这些 a 都看作已知的, 我们感兴趣的是变量 $x_1,\cdots,x_n$ 的解. 因此, 我们有 m 个方程和 n 个变量. 显然 $x_1=\cdots=x_n=0$ 是一个解, 关键问题是, 是否还有其他解. 也就是说, 我们想知道 $\operatorname{null}T$ 是否严格大于 $\{\boldsymbol{0}\}$. 这种情形恰好当 T 不是单的时才出现 (由 3.2). 由 3.5 可知, 如果 $n>m$, 那么 T 不是单的. 结论: 当变量多于方程时, 齐次线性方程组必有非零解.

设 T 如上一段, 现在考虑方程 $T\boldsymbol{x}=\boldsymbol{c}$, 其中 $\boldsymbol{c}=(c_1,\cdots,c_m)\in\mathbf{F}^m$. 把方程 $T\boldsymbol{x}=\boldsymbol{c}$ 重新写成一个非齐次方程组:

变量多于方程的齐次方程组和方程多于变量的非齐次方程组的这些结果通常都是用高斯消元法来证明的. 这里所采取的抽象的处理方法使证明更简洁.

$$\begin{gathered}\sum_{k=1}^{n}a_{1,k}x_k=c_1\\ \vdots\\ \sum_{k=1}^{n}a_{m,k}x_k=c_m.\end{gathered}$$

和以前一样, 把其中的这些 a 都看作已知的. 现在的关键问题是, 是不是对于每一组常数项 $c_1,\cdots,c_m\in\mathbf{F}$, 变量 $x_1,\cdots,x_n$ 都至少有一个解. 也就是说, 我们想知道 $\operatorname{range}T$ 是否等于 $\mathbf{F}^m$. 由 3.6 可知, 如果 $n<m$, 那么 T 不是满的. 结论: 当方程多于变量时, 必有一组常数项使得相应的非齐次线性方程组无解.

§3.3 线性映射的矩阵

我们已经看到, 如果 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基, 并且 $T:V\to W$ 是线性的, 那么值 $T\boldsymbol{v}_1,\cdots,T\boldsymbol{v}_n$ 确定了 T 在 V 中任意向量上的值. 在这一节, 我们将看到如何用矩阵和 W 的基来有效地记录这些 $T\boldsymbol{v}_j$ 的值.

设 m 和 n 都是正整数. 一个 $m\times n$ **矩阵 (matrix)** 是一个有 m 个行和 n 个列的矩形阵列, 犹如:

3.7
$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}.$$

注意, 第一个下标代表行数, 第二个下标代表列数. 因此, $a_{3,2}$ 表明该元素位于上述矩阵的第三行第二列. 我们考虑的矩阵的元素通常都是 $\mathbf{F}$ 中的元素.

设 $T\in\mathcal{L}(V,W)$, $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基, $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 是 W 的基, 那么对于每个 $k=1,\cdots,n$, $T\boldsymbol{v}_k$ 都可以唯一地写成这些 $\boldsymbol{w}$ 的线性组合:

3.8
$$T\boldsymbol{v}_k=a_{1,k}\boldsymbol{w}_1+\cdots+a_{m,k}\boldsymbol{w}_m,$$

其中 $a_{j,k}\in\mathbf{F}$, $j=1,\cdots,m$. 因为线性映射由其在基上的值确定, 所以线性映射 T 由这些标量 $a_{j,k}$ 完全确定. 由这些 a 所构成的 $m\times n$ 矩阵 3.7 称为 T 关于基 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 和基 $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 的**矩阵 (matrix)**, 记为

$$\mathcal{M}(T,(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n),(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)).$$

如果基 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 和基 $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 在上下文中是自明的 (例如, 只有一个基), 那么将 $\mathcal{M}(T,(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n),(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m))$ 简记为 $\mathcal{M}(T)$.

为了记住如何从 T 构造 $\mathcal{M}(T)$, 可以将定义域的基向量 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 横写在顶端, 并将目标空间的基向量 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m$ 竖

写在左侧, 如下所示:

$$\begin{array}{c} \\ \boldsymbol{w}_1 \\ \vdots \\ \boldsymbol{w}_m \end{array}\begin{array}{c} \boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_k \ \cdots \ \boldsymbol{v}_n \\ \left[\begin{array}{ccc} & a_{1,k} & \\ & \vdots & \\ & a_{m,k} & \end{array}\right] \end{array}.$$

注意, 在上面的矩阵中, 只列出了第 k 列 (因此所列出的这些 a 的第二个下标都是 k). 把 $T\boldsymbol{v}_k$ 写成诸 $\boldsymbol{w}$ 的线性组合, 所需的这些系数就组成了 $\mathcal{M}(T)$ 的第 k 列. 于是, 上面的图会提醒你, 从矩阵 $\mathcal{M}(T)$ 又得到了 $T\boldsymbol{v}_k$: 将矩阵的第 k 列的每个元素与左侧的列中相应的 $\boldsymbol{w}$ 相乘, 然后再将所得向量相加.

线性映射 0 (把每个向量都变成 0 的线性映射)关于任何基的矩阵都由 0 组成.

如果 T 是 $\mathbf{F}^n$ 到 $\mathbf{F}^m$ 的线性映射, 那么除非特殊说明, 总设所考虑的基是标准基 (其中第 k 个基向量的第 k 个位置是 1, 而其他位置都是 0). 如果把 $\mathbf{F}^m$ 中的元素看成由 m 个数组成的列, 那么可以把 $\mathcal{M}(T)$ 的第 k 列视为 T 对 k 个基向量的作用. 例如, 若 $T \in \mathcal{L}(\mathbf{F}^2, \mathbf{F}^3)$ 定义如下

$$T(x, y) = (x + 3y, 2x + 5y, 7x + 9y),$$

那么 $T(1,0) = (1,2,7)$, $T(0,1) = (3,5,9)$. 因此, T (关于标准基) 的矩阵是 3×2 矩阵

$$\begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{bmatrix}.$$

若我们有 V 的基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 和 W 的基 $(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m)$, 则对于每个从 V 到 W 的线性映射, 我们都可以谈论它的矩阵 (当然, 是关于这些基的). 两个线性映射之和的矩阵是否等于这两个映射的矩阵之和?

这个问题现在还没有意义, 这是因为, 尽管我们定义了两个线性映射的和, 但是还没有定义两个矩阵的和. 幸好矩阵的和有明显的定义, 而且此定义恰好就有这样的性质. 具体地, 我们规定同样大小的矩阵相加就是把矩阵中相对应的元素相加:

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} + \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{bmatrix}$$
$$= \begin{bmatrix} a_{1,1}+b_{1,1} & \cdots & a_{1,n}+b_{1,n} \\ \vdots & & \vdots \\ a_{m,1}+b_{m,1} & \cdots & a_{m,n}+b_{m,n} \end{bmatrix}.$$

对于如此定义的矩阵加法, 你应该验证, 当 $T, S \in \mathcal{L}(V, W)$ 时,

3.9 $$\mathcal{M}(T+S) = \mathcal{M}(T) + \mathcal{M}(S).$$

仍然假设我们已经取定了某些基, 标量与线性映射之积的矩阵是否等于标量与线性映射的矩阵之积? 这个问题现在还没有意义, 这是因为我们还没有定义矩阵的标量乘法. 幸好矩阵的标量乘法有明显的定义, 而且此定义恰好就有这样的性质. 具体地, 标量与矩阵的乘积就是用该标量乘以矩阵的每个元素:

$$c\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} = \begin{bmatrix} ca_{1,1} & \cdots & ca_{1,n} \\ \vdots & & \vdots \\ ca_{m,1} & \cdots & ca_{m,n} \end{bmatrix}.$$

对于如此定义的矩阵的标量乘法, 你应该验证, 当 $c \in \mathbf{F}$, $T \in \mathcal{L}(V, W)$ 时,

3.10 $$\mathcal{M}(cT) = c\mathcal{M}(T).$$

因为已经定义了矩阵的加法和标量乘法, 所以由此产生一个向量空间就不足为奇了, 只需一个记号来表示这个向量空间, 元素在 $\mathbf{F}$ 中的所有 $m \times n$ 矩阵之集记为 $\mathrm{Mat}(m, n, \mathbf{F})$. 你应该验证, 关于上面定义的矩阵加法和标量乘法, $\mathrm{Mat}(m, n, \mathbf{F})$ 是向量空间. 注意, $\mathrm{Mat}(m, n, \mathbf{F})$ 的加法单位元是元素全为 0 的 $m \times n$ 矩阵.

设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, $(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m)$ 是 W 的基, $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p)$ 是 另一个向量空间 U 的基. 考虑线性映射 $S : U \to V$ 和

$T: V \to W$, 它们的复合映射 TS 是从 U 到 W 的线性映射. 如何通过 $\mathcal{M}(T)$ 和 $\mathcal{M}(S)$ 计算 $\mathcal{M}(TS)$? 这个问题的最佳答案是如下的漂亮关系:

3.11 $$\mathcal{M}(TS) = \mathcal{M}(T)\mathcal{M}(S).$$

但是, 这个方程的右端目前还没有意义, 这是因为我们还没有定义矩阵的乘法. 我们要定义一种矩阵乘法以使上式成立. 让我们来看看该怎样做.

设

$$\mathcal{M}(T) = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}, \qquad \mathcal{M}(S) = \begin{bmatrix} b_{1,1} & \cdots & b_{1,p} \\ \vdots & & \vdots \\ b_{n,1} & \cdots & b_{n,p} \end{bmatrix}.$$

对于 $k \in \{1, \cdots, p\}$, 我们有

$$\begin{aligned} TS\boldsymbol{u}_k &= T\left(\sum_{r=1}^{n} b_{r,k}\boldsymbol{v}_r\right) \\ &= \sum_{r=1}^{n} b_{r,k}T\boldsymbol{v}_r \\ &= \sum_{r=1}^{n} b_{r,k}\sum_{j=1}^{m} a_{j,r}\boldsymbol{w}_j \\ &= \sum_{j=1}^{m}\left(\sum_{r=1}^{n} a_{j,r}b_{r,k}\right)\boldsymbol{w}_j. \end{aligned}$$

因此, $\mathcal{M}(TS)$ 是 $m \times p$ 矩阵, 它的第 j 行第 k 列元素等于 $\sum_{r=1}^{n} a_{j,r}b_{r,k}$.

现在如何定义矩阵的乘法以使得 3.11 成立就是显然的事了. 即, 如果 $\boldsymbol{A}$ 是 $m \times n$ 矩阵, 其元素为 $a_{j,k}$, $\boldsymbol{B}$ 是 $n \times p$ 矩阵, 其元素为 $b_{j,k}$, 那么定义 $\boldsymbol{AB}$ 是 $m \times p$ 矩阵, 其第 j 行第 k 列元素等于

你可能从以前的课程已经知道了矩阵乘法的定义, 尽管当时你可能还不知道这么做的动机.

$$\sum_{r=1}^{n} a_{j,r}b_{r,k}.$$

也就是说, 把 $\boldsymbol{A}$ 的第 j 行与 $\boldsymbol{B}$ 的第 k 列的对应元素相乘再求和, 就得到 $\boldsymbol{AB}$ 中第 j 行第 k 列元素. 注意, 只有当第一个矩阵

你应该举例说明矩阵乘法是不可交换的. 也就是说, $\boldsymbol{AB}$ 不一定等于 $\boldsymbol{BA}$, 即使它们都有定义.

的列数等于第二个矩阵的行数时, 我们才能定义这两个矩阵的乘积.

例如, 3×2 矩阵乘以 2×4 矩阵可得一个 3×4 矩阵:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}\begin{bmatrix} 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & -1 \end{bmatrix}=\begin{bmatrix} 10 & 7 & 4 & 1 \\ 26 & 19 & 12 & 5 \\ 42 & 31 & 20 & 9 \end{bmatrix}.$$

设 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基. 如果 $\boldsymbol{v}\in V$, 那么存在唯一一组数 $b_1,\cdots,b_n$, 使得

3.12
$$\boldsymbol{v}=b_1\boldsymbol{v}_1+\cdots+b_n\boldsymbol{v}_n.$$

$\boldsymbol{v}$ 的**矩阵 (matrix)** 是一个 $n\times 1$ 矩阵, 记为 $\mathcal{M}(\boldsymbol{v})$, 定义如下

3.13
$$\mathcal{M}(\boldsymbol{v})=\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

在上下文中, 基通常都是自明的, 但是, 当采用符号 $\mathcal{M}(\boldsymbol{v},(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n))$ 而不采用 $\mathcal{M}(\boldsymbol{v})$ 时, 就需要把基明确地写出来.

例如, 把向量 $\boldsymbol{x}\in\mathbf{F}^n$ 的坐标看成一个 $n\times 1$ 矩阵中的元素, 即可得 $\boldsymbol{x}$ 关于标准基的矩阵. 也就是说, 如果 $\boldsymbol{x}=(x_1,\cdots,x_n)\in\mathbf{F}^n$, 那么

$$\mathcal{M}(\boldsymbol{x})=\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

下一个命题表明, 线性映射的矩阵、向量的矩阵以及矩阵乘法是如何结合在一起的. 在此命题中, $\mathcal{M}(T\boldsymbol{v})$ 是向量 $T\boldsymbol{v}$ 关于基 $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 的矩阵, $\mathcal{M}(\boldsymbol{v})$ 是向量 $\boldsymbol{v}$ 关于基 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 的矩阵, 而 $\mathcal{M}(T)$ 是线性映射 T 关于基 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 和 $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 的矩阵.

3.14 命题: 设 $T\in\mathcal{L}(V,W)$, $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基, $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 是 W 的基, 那么对每个 $\boldsymbol{v}\in V$ 都有

$$\mathcal{M}(T\boldsymbol{v})=\mathcal{M}(T)\mathcal{M}(\boldsymbol{v}).$$

证明：设

3.15
$$\mathcal{M}(T)=\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}$$

回想一下, 这指的是, 对每个 k 都有,

3.16
$$T\boldsymbol{v}_k=\sum_{j=1}^{m}a_{j,k}\boldsymbol{w}_j$$

设 $\boldsymbol{v}$ 是 V 中的任意向量, 将其写成 3.12 的形式, 那么 $\mathcal{M}(\boldsymbol{v})$ 就是 3.13 所给出的矩阵. 现在,

$$\begin{aligned} T\boldsymbol{v}&=b_1T\boldsymbol{v}_1+\cdots+b_nT\boldsymbol{v}_n \\ &=b_1\sum_{j=1}^{m}a_{j,1}\boldsymbol{w}_j+\cdots+b_n\sum_{j=1}^{m}a_{j,n}\boldsymbol{w}_j \\ &=\sum_{j=1}^{m}(a_{j,1}b_1+\cdots+a_{j,n}b_n)\boldsymbol{w}_j, \end{aligned}$$

其中第一个等式由 3.12 得到, 第二个等式由 3.16 得到. 最后的那个等式表明, 向量 $T\boldsymbol{v}$ 关于基 $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 的 $m\times 1$ 矩阵 $\mathcal{M}(T\boldsymbol{v})$ 由下面的等式给出,

$$\mathcal{M}(T\boldsymbol{v})=\begin{bmatrix} a_{1,1}b_1+\cdots+a_{1,n}b_n \\ \vdots \\ a_{m,1}b_1+\cdots+a_{m,n}b_n \end{bmatrix}.$$

根据公式 3.15 和 3.13 以及矩阵乘法的定义, 这个公式表明 $\mathcal{M}(T\boldsymbol{v})=\mathcal{M}(T)\mathcal{M}(\boldsymbol{v})$. ■

§3.4 可 逆 性

线性映射 $T\in\mathcal{L}(V,W)$ 称为 **可逆的 (invertible)**, 如果存在线性映射 $S\in\mathcal{L}(W,V)$ 使得 ST 等于 V 上的恒等映射, 并且 TS 等于 W 上的恒等映射. 满足 $ST=I$, $TS=I$ 的线性映射 $S\in\mathcal{L}(W,V)$ 称为 T 的 **逆 (inverse)** (注意, 第一个 I 是 V 上的恒等映射, 第二个 I 是 W 上的恒等映射).

如果 S 和 S' 都是 T 的逆, 那么

$$S = SI = S(TS') = (ST)S' = IS' = S',$$

故 $S = S'$. 也就是说, 如果 T 可逆, 那么它的逆是唯一的, 将其记为 T^{-1}. 再重复一遍: 如果 $T \in \mathcal{L}(V, W)$ 可逆, 那么 T^{-1} 是 $\mathcal{L}(W, V)$ 中唯一一个使得 $T^{-1}T = I$, $TT^{-1} = I$ 的元素. 下面的命题刻划了可逆线性映射.

3.17 命题： 一个线性映射是可逆的当且仅当它既是单的又是满的.

证明： 设 $T \in \mathcal{L}(V, W)$, 那么需要证明, T 是可逆的当且仅当它既是单的又是满的.

首先假设 T 是可逆的. 为了证明 T 是单的, 设 $\boldsymbol{u}, \boldsymbol{v} \in V$, $T\boldsymbol{u} = T\boldsymbol{v}$, 那么

$$\boldsymbol{u} = T^{-1}(T\boldsymbol{u}) = T^{-1}(T\boldsymbol{v}) = \boldsymbol{v},$$

故 $\boldsymbol{u} = \boldsymbol{v}$. 因此, T 是单的.

仍设 T 是可逆的. 现在证明 T 是满的. 为此, 设 $\boldsymbol{w} \in W$, 那么 $\boldsymbol{w} = T(T^{-1}\boldsymbol{w})$, 这表明 $\boldsymbol{w}$ 含于 T 的值域. 于是 $\operatorname{range} T = W$, 故 T 是满的. 这就完成了证明的一个方面.

现在假设 T 既是单的又是满的. 我们需要证明 T 是可逆的. 对于每个 $\boldsymbol{w} \in W$, 定义 $S\boldsymbol{w}$ 是 V 中唯一使得 $T(S\boldsymbol{w}) = \boldsymbol{w}$ 的那个元素 (由 T 既单又满可得此元素的存在性和唯一性). 显然, TS 等于 W 上的恒等映射. 为了证明 ST 等于 V 上的恒等映射, 设 $\boldsymbol{v} \in V$, 那么

$$T(ST\boldsymbol{v}) = (TS)(T\boldsymbol{v}) = I(T\boldsymbol{v}) = T\boldsymbol{v}.$$

这个等式表明 $ST\boldsymbol{v} = \boldsymbol{v}$ (因 T 是单的), 因此 ST 等于 V 上的恒等映射. 为了完成证明, 还需要证明 S 是线性的. 为此, 设 $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W$, 那么

$$T(S\boldsymbol{w}_1 + S\boldsymbol{w}_2) = T(S\boldsymbol{w}_1) + T(S\boldsymbol{w}_2) = \boldsymbol{w}_1 + \boldsymbol{w}_2.$$

于是, $S\boldsymbol{w}_1 + S\boldsymbol{w}_2$ 是 V 中唯一被 T 映成 $\boldsymbol{w}_1 + \boldsymbol{w}_2$ 的那个元素. 再由 S 的定义可得 $S(\boldsymbol{w}_1 + \boldsymbol{w}_2) = S\boldsymbol{w}_1 + S\boldsymbol{w}_2$. 因此, S 满足加性. 齐性的证明是类似的. 具体地, 如果 $\boldsymbol{w} \in W$, $a \in \mathbf{F}$, 那么

$$T(aS\boldsymbol{w}) = aT(S\boldsymbol{w}) = a\boldsymbol{w}.$$

于是, $aS\boldsymbol{w}$ 是 V 中唯一被 T 映成 $a\boldsymbol{w}$ 的那个元素. 再由 S 的定义可得 $S(a\boldsymbol{w}) = aS\boldsymbol{w}$. 因此, S 是线性的. ■

称两个向量空间是**同构的 (isomorphic)**, 如果存在从一个向量空间到另一个向量空间的可逆线性映射. 作为抽象的向量空间, 两个同构的向量空间具有相同的性质. 按照这种观点, 你可以认为可逆线性映射就是把向量空间的元素都重新标记了一遍.

在希腊语中, 词 isos 的意思是相同; 词 morph 的意思是形状. 因此, isomorphic 的字面意思就是同形.

如果两个向量空间是同构的, 并且其中之一是有限维的, 那么另一个也是有限维的. 为了说明这一点, 设 V 与 W 是同构的, 并且 $T \in \mathcal{L}(V, W)$ 是可逆线性映射. 若 V 是有限维的, 则 W 亦然 (由 3.4). 用 $T^{-1} \in \mathcal{L}(W, V)$ 代替 T 做同样的推理可以证明, 若 W 是有限维的, 则 V 亦然. 实际上, 有更多的事实成立, 如下面的定理所示.

3.18 定理: 两个有限维向量空间同构当且仅当它们的维数相等.

证明: 首先设 V 和 W 是同构的有限维向量空间, 则存在从 V 到 W 的可逆线性映射 T. 因为 T 是可逆的, 所以 $\operatorname{null} T = \{\boldsymbol{0}\}$, $\operatorname{range} T = W$, 从而, $\dim \operatorname{null} T = 0$, $\dim \operatorname{range} T = \dim W$. 于是, 公式

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$$

(参见 3.4) 变成了等式 $\dim V = \dim W$, 这就证明了一个方面.

为了证明另一个方面, 假设 V 和 W 是维数相同的向量空间, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, $(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$ 是 W 的基, T 是从 V 到 W 的线性映射, 定义如下

$$T(a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n) = a_1\boldsymbol{w}_1 + \cdots + a_n\boldsymbol{w}_n.$$

因为 $(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$ 张成 W, 所以 T 是满的; 又因为 $(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n)$ 是线性无关的, 所以 T 还是单的. 由于 T 既单又满, 从而可逆 (参见 3.17), 因此 V 与 W 是同构的. ■

既然每个有限维向量空间都同构于某个 $\mathbf{F}^n$，那为什么还要研究抽象的向量空间呢？答案是，$\mathbf{F}^n$ 的研究会很快得出不等于 $\mathbf{F}^n$ 的向量空间。例如，我们会遇到线性映射的零空间和值域、矩阵集合 $\operatorname{Mat}(n,n,\mathbf{F})$、多项式集合 $\mathcal{P}_n(\mathbf{F})$。尽管这些向量空间都分别同构于某个 $\mathbf{F}^m$，但是这样考虑问题往往只会增加复杂性而不会带来新的见识。

上面的这个定理表明，每个有限维向量空间都同构于某个 $\mathbf{F}^n$. 具体地，如果 V 是有限维向量空间，且 $\dim V = n$，那么 V 和 $\mathbf{F}^n$ 同构.

如果 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基，$(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 是 W 的基，那么对于每个 $T\in\mathcal{L}(V,W)$，都有矩阵 $\mathcal{M}(T)\in\operatorname{Mat}(m,n,\mathbf{F})$. 这就是说，一旦选定了 V 和 W 的基，那么 $\mathcal{M}$ 就是从 $\mathcal{L}(V,W)$ 到 $\operatorname{Mat}(m,n,\mathbf{F})$ 的函数. 注意，3.9 和 3.10 表明 $\mathcal{M}$ 是线性映射. 现在我们证明，这个线性映射实际上还是可逆的.

3.19 命题： 设 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基，$(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m)$ 是 W 的基，那么 $\mathcal{M}$ 是 $\mathcal{L}(V,W)$ 和 $\operatorname{Mat}(m,n,\mathbf{F})$ 之间的可逆线性映射.

证明： 已知 $\mathcal{M}$ 是线性的，故只需证明 $\mathcal{M}$ 既是单的又是满的 (由 3.17). 这些都很容易. 先从单性开始. 如果 $T\in\mathcal{L}(V,W)$，并且 $\mathcal{M}(T)=\mathbf{0}$，那么 $T\boldsymbol{v}_k=\mathbf{0}$，$k=1,\cdots,n$. 因为 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基，所以 $T=0$. 于是，$\mathcal{M}$ 是单的 (由 3.2).

为了证明 $\mathcal{M}$ 是满的，设

$$\boldsymbol{A}=\begin{bmatrix} a_{1,1} & \cdots & a_{1,n}\\ \vdots & & \vdots\\ a_{m,1} & \cdots & a_{m,n}\end{bmatrix}$$

是 $\operatorname{Mat}(m,n,\mathbf{F})$ 中的一个矩阵，T 是从 V 到 W 的线性映射，使得

$$T\boldsymbol{v}_k=\sum_{j=1}^{m} a_{j,k}\boldsymbol{w}_j,\quad k=1,\cdots,n.$$

显然，$\mathcal{M}(T)$ 等于 $\boldsymbol{A}$，故 $\mathcal{M}$ 的值域等于 $\operatorname{Mat}(m,n,\mathbf{F})$. ■

显然，只有一个元素为 1，其余元素全为 0 的 $m\times n$ 矩阵的全体组成了 $\operatorname{Mat}(m,n,\mathbf{F})$ 的基. 这样的矩阵共有 mn 个，因此 $\operatorname{Mat}(m,n,\mathbf{F})$ 的维数等于 mn.

现在我们可以确定从一个向量空间到另一个向量空间的所有线性映射所构成的向量空间的维数.

3.20 命题: 如果 V 和 W 都是有限维的, 那么 $\mathcal{L}(V,W)$ 是有限维的, 并且

$$\dim \mathcal{L}(V,W) = (\dim V)(\dim W).$$

证明: 由 $\dim \operatorname{Mat}(m,n,\mathbf{F}) = mn$, 3.18 以及 3.19 得证. ■

一个向量空间到其自身的线性映射称为**算子 (operator)**. 如果想指明向量空间, 我们说线性映射 $T: V \to V$ 是 V 上的算子. 因为我们感兴趣的常常是从一个向量空间到其自身的线性映射, 所以我们采用符号 $\mathcal{L}(V)$ 来表示 V 上算子的集合. 也就是说, $\mathcal{L}(V) = \mathcal{L}(V,V)$.

算子的研究是线性代数中, 也是本书中最深刻和最重要的内容.

由 3.17 可知, 若一个线性映射既单又满, 则其可逆. 对于向量空间到其自身的线性映射, 你可能想知道, 仅由单性或满性之一是否可以推出可逆性. 从我们考虑过的某些例子可以看出, 对于无限维向量空间, 任何一个条件都不能单独推出可逆性. x^2 乘算子 (从 $\mathcal{P}(\mathbf{R})$ 到其自身) 是单的但不是满的. 后向移位算子 (从 $\mathbf{F}^\infty$ 到其自身) 是满的但不是单的. 从这些例子来看, 下一个定理是很不平凡的 —— 它表明, 对于从有限维向量空间到其自身的线性映射, 单性和满性中的任何一个都能推出另一个.

3.21 定理: 设 V 是有限维的. 如果 $T \in \mathcal{L}(V)$, 那么下列等价:

(a) T 是可逆的;

(b) T 是单的;

(c) T 是满的.

证明: 设 $T \in \mathcal{L}(V)$. 显然, (a) 蕴含 (b).

现在假设 (b) 成立, 于是 T 是单的, 从而 $\operatorname{null} T = \mathbf{0}$ (由 3.2). 由 3.4 可得

$$\begin{aligned}\dim \operatorname{range} T &= \dim V - \dim \operatorname{null} T \\ &= \dim V,\end{aligned}$$

这表明 $\operatorname{range} T$ 等于 V (参见第 2 章习题 11), 从而 T 是满的. 因此 (b) 蕴含 (c).

现在假设 (c) 成立, 于是 T 是满的, 从而 $\operatorname{range} T = V$. 由

3.4 可得

$$\begin{aligned}\dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T \\ &= 0,\end{aligned}$$

这表明 null T 等于 $\{\mathbf{0}\}$, 从而 T 是单的 (由 3.2), 故 T 是可逆的 (已知 T 是满的). 因此 (c) 蕴含 (a). ■

习　题

1. 证明每个从一维向量空间到其自身的线性映射都是乘以某个标量. 准确地说, 证明: 如果 $\dim V = 1$, $T \in \mathcal{L}(V,V)$, 那么有 $a \in \mathbf{F}$ 使得对所有 $\boldsymbol{v} \in V$ 都有 $T\boldsymbol{v} = a\boldsymbol{v}$.

2. 给出一个非线性函数 $f: \mathbf{R}^2 \to \mathbf{R}$ 使得对所有 $a \in \mathbf{R}$ 和所有 $\boldsymbol{v} \in \mathbf{R}^2$ 都有

$$f(a\boldsymbol{v}) = af(\boldsymbol{v}).$$

习题 2 表明齐性并不意味着线性. 加性也不意味着线性, 其证明涉及的高等工具超出了本书的范围.

3. 设 V 是有限维的. 证明 V 的子空间上的线性映射可以扩张成 V 上的线性映射. 也就是说, 证明: 如果 U 是 V 的子空间, $S \in \mathcal{L}(U,W)$, 那么存在 $T \in \mathcal{L}(V,W)$, 使得对所有 $\boldsymbol{u} \in U$ 都有 $T\boldsymbol{u} = S\boldsymbol{u}$.

4. 设 T 是从 V 到 $\mathbf{F}$ 的线性映射. 证明: 若 $\boldsymbol{u} \in V$ 不含于 $\operatorname{null} T$, 则

$$V = \operatorname{null} T \oplus \{a\boldsymbol{u} : a \in \mathbf{F}\}.$$

5. 设 $T \in \mathcal{L}(V,W)$ 是单的, 并且 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 在 V 中线性无关. 证明 $(T\boldsymbol{v}_1, \cdots, T\boldsymbol{v}_n)$ 在 W 中线性无关.

6. 证明: 如果 $S_1, \cdots, S_n$ 都是单的线性映射, 并且 $S_1 \cdots S_n$ 有意义, 那么 $S_1 \cdots S_n$ 是单的.

7. 证明: 如果 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 张成 V, 并且 $T \in \mathcal{L}(V,W)$ 是满的, 那么 $(T\boldsymbol{v}_1, \cdots, T\boldsymbol{v}_n)$ 张成 W.

8. 设 V 是有限维的, 并且 $T \in \mathcal{L}(V,W)$. 证明 V 有子空间 U 使得 $U \cap \operatorname{null} T = \{\mathbf{0}\}$, 并且 $\operatorname{range} T = \{T\boldsymbol{u} : \boldsymbol{u} \in U\}$.

9. 证明: 如果 T 是从 $\mathbf{F}^4$ 到 $\mathbf{F}^2$ 的线性映射, 使得

$$\operatorname{null} T = \{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_1 = 5x_2 \text{ 且 } x_3 = 7x_4\},$$

那么 T 是满的.

10. 证明从 $\mathbf{F}^5$ 到 $\mathbf{F}^2$ 的线性映射的零空间都不等于

$$\{(x_1, x_2, x_3, x_4, x_5) \in \mathbf{F}^5 : x_1 = 3x_2 \text{ 且 } x_3 = x_4 = x_5\}.$$

11. 证明：如果在 V 上有一个线性映射，其零空间和值域都是有限维的，那么 V 是有限维的.

12. 设 V 和 W 都是有限维的. 证明存在从 V 到 W 的满的线性映射当且仅当 $\dim W \leqslant \dim V$.

13. 设 V 和 W 都是有限维的，并且 U 是 V 的子空间. 证明：存在 $T \in \mathcal{L}(V, W)$ 使得 $\operatorname{null} T = U$ 当且仅当 $\dim U \geqslant \dim V - \dim W$.

14. 设 W 是有限维的，并且 $T \in \mathcal{L}(V, W)$. 证明：T 是单的当且仅当有 $S \in \mathcal{L}(W, V)$ 使得 ST 是 V 上的恒等映射.

15. 设 V 是有限维的，并且 $T \in \mathcal{L}(V, W)$. 证明：T 是满的当且仅当有 $S \in \mathcal{L}(W, V)$ 使得 TS 是 W 上的恒等映射.

16. 设 U 和 V 都是有限维向量空间，并且 $S \in \mathcal{L}(V, W)$，$T \in \mathcal{L}(U, V)$. 证明

$$\dim \operatorname{null} ST \leqslant \dim \operatorname{null} S + \dim \operatorname{null} T.$$

17. 证明矩阵加法和乘法的分配性质成立. 也就是说，设 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 都是矩阵，并且 $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C})$ 有意义. 证明：$\boldsymbol{AB}+\boldsymbol{AC}$ 有意义，并且 $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}$.

18. 证明矩阵乘法是结合的. 也就是说，设 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 都是矩阵，并且 $(\boldsymbol{AB})\boldsymbol{C}$ 有意义. 证明：$\boldsymbol{A}(\boldsymbol{BC})$ 有意义，并且 $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$.

这个习题表明 T 具有在第 39 页所保证的形式.

19. 设 $T \in \mathcal{L}(\mathbf{F}^n, \mathbf{F}^m)$，并且

$$\mathcal{M}(T) = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix},$$

其中使用了标准基. 证明: 对于每个 $(x_1,\cdots,x_n)\in\mathbf{F}^n$ 都有

$$T(x_1,\cdots,x_n)=(a_{1,1}x_1+\cdots+a_{1,n}x_n,\cdots,\\ a_{m,1}x_1+\cdots+a_{m,n}x_n).$$

20. 设 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基. 证明如下定义的函数 $T:V\to\mathrm{Mat}(n,1,\mathbf{F})$,

$$T\boldsymbol{v}=\mathcal{M}(\boldsymbol{v}),$$

是 V 到 $\mathrm{Mat}(n,1,\mathbf{F})$ 上的可逆线性映射, 其中 $\mathcal{M}(\boldsymbol{v})$ 是 $\boldsymbol{v}\in V$ 关于基 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 的矩阵.

21. 证明: 从 $\mathrm{Mat}(n,1,\mathbf{F})$ 到 $\mathrm{Mat}(m,1,\mathbf{F})$ 的每个线性映射都是乘以某个矩阵. 换句话说, 证明: 如果 $T\in\mathcal{L}(\mathrm{Mat}(n,1,\mathbf{F}),\mathrm{Mat}(m,1,\mathbf{F}))$, 那么有 $m\times n$ 矩阵 $\boldsymbol{A}$ 使得对每个 $\boldsymbol{B}\in\mathrm{Mat}(n,1,\mathbf{F})$ 都有 $T\boldsymbol{B}=\boldsymbol{AB}$.

22. 设 V 是有限维的, 并且 $S,T\in\mathcal{L}(V)$. 证明 ST 可逆当且仅当 S 和 T 都可逆.

23. 设 V 是有限维的, 并且 $S,T\in\mathcal{L}(V)$. 证明 $ST=I$ 当且仅当 $TS=I$.

24. 设 V 是有限维的, 并且 $T\in\mathcal{L}(V)$. 证明 T 是恒等映射的标量倍当且仅当对每个 $S\in\mathcal{L}(V)$ 都有 $ST=TS$.

25. 证明: 如果 V 是有限维的, 并且 $\dim V>1$, 那么 V 上不可逆算子之集不是 $\mathcal{L}(V)$ 的子空间.

26. 设 n 是正整数, 并且 $a_{i,j}\in\mathbf{F}$, $i,j=1,\cdots,n$. 证明下面的 (a) 和 (b) 等价.

    (a) 齐次线性方程组

$$\sum_{k=1}^{n}a_{1,k}x_k=0$$

$$\vdots$$

$$\sum_{k=1}^{n}a_{n,k}x_k=0.$$

只有平凡解 $x_1 = \cdots = x_n = 0$.

(b) 对于每组 $c_1, \cdots, c_n \in \mathbf{F}$, 方程组

$$\sum_{k=1}^{n} a_{1,k} x_k = c_1$$
$$\vdots$$
$$\sum_{k=1}^{n} a_{n,k} x_k = c_n.$$

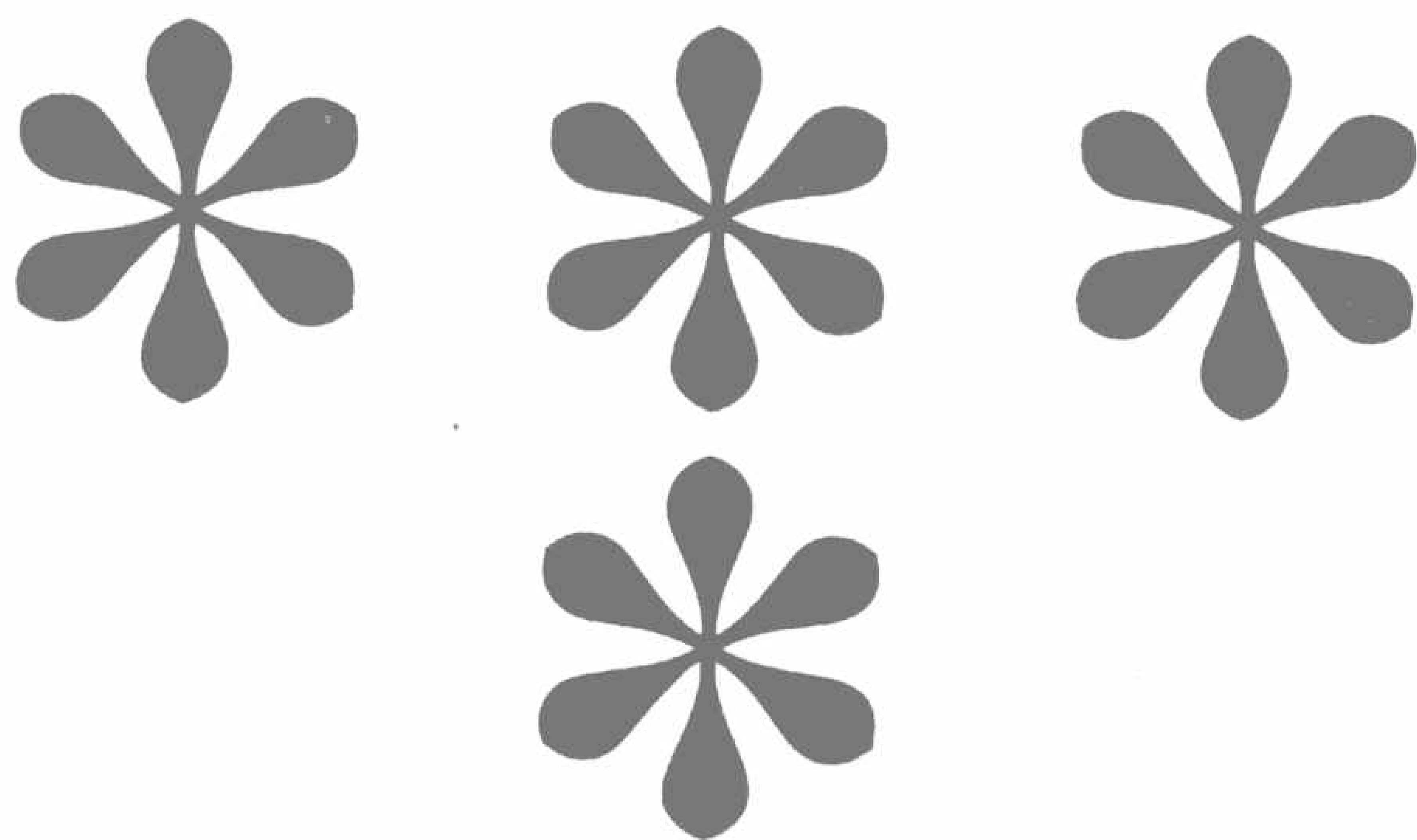
都有解.

注意, 此处方程的个数与变量的个数相同.

第 4 章　多　项　式

这简短的一章不包含线性代数的内容, 而是包含了我们在讨论从向量空间到其自身的线性映射时所需要的关于多项式的一些背景资料. 本章的很多结果你可能已经从其他课程知道了; 在这里包含这些知识只是为了完整性. 因为这一章不是关于线性代数的, 所以老师可能会讲得很快. 可以不要求你把所有的证明都仔细看一遍, 但是你至少要把本章中所有结论的陈述看一遍, 并且理解它们 —— 这些结论在本书的其余部分是要用到的.

F 表示 **R** 或 **C**.



§4.1 次 数

回忆一下, 对于函数 $p:\mathbf{F}\to\mathbf{F}$, 如果存在 $a_0,\cdots,a_m\in\mathbf{F}$ 使得对所有 $z\in\mathbf{F}$ 都有

$$p(z)=a_0+a_1z+a_2z^2+\cdots+a_mz^m,$$

则称 p 为系数在 $\mathbf{F}$ 中的多项式. 如果 p 可以写成上述形式, 其中 $a_m\neq 0$, 则称 p 的次数为 m. 如果所有的系数 $a_0,\cdots,a_m$ 都等于 0, 那么我们就说 p 的次数为 $-\infty$. 就目前我们所知道的而言, 一个多项式可能有不止一个次数, 因为我们还没有证明上面等式中的系数是由函数 p 唯一确定的.

必要的时候, 可以使用关于 $-\infty$ 的显然算术. 比如, 对每个整数 m, 都有 $-\infty<m$, 并且 $-\infty+m=-\infty$. 我们已经规定了 0 多项式的次数是 $-\infty$, 于是, 很多结果都不再有例外. 例如, 即使 $p=0$, pq 的次数也等于 p 的次数加上 q 的次数.

回忆一下, $\mathcal{P}(\mathbf{F})$ 表示系数在 $\mathbf{F}$ 中的所有多项式所组成的向量空间, $\mathcal{P}_m(\mathbf{F})$ 是 $\mathcal{P}(\mathbf{F})$ 的子空间, 由系数在 $\mathbf{F}$ 中次数不超过 m 的多项式组成. 对于多项式 $p\in\mathcal{P}(\mathbf{F})$, 如果数 $\lambda\in\mathbf{F}$ 满足

$$p(\lambda)=0.$$

则称 λ 为 p 的**根 (root)**. 根在多项式的研究中起着至关重要的作用. 我们首先证明, λ 是 p 的根当且仅当 p 是 $z-\lambda$ 的倍式.

4.1 命题: *设 $p\in\mathcal{P}(\mathbf{F})$ 是 m 次多项式, $m\geqslant 1$. 令 $\lambda\in\mathbf{F}$, 则 λ 是 p 的根当且仅当存在 $m-1$ 次多项式 $q\in\mathcal{P}(\mathbf{F})$ 使得*

4.2 $$p(z)=(z-\lambda)q(z),\quad z\in\mathbf{F}.$$

证明: 证明的一个方面是显然的. 即假设存在多项式 $q\in\mathcal{P}(\mathbf{F})$ 使得 4.2 成立, 则

$$p(\lambda)=(\lambda-\lambda)q(\lambda)=0,$$

因此 λ 是 p 的根.

要证明另一个方面, 设 $\lambda\in\mathbf{F}$ 是 p 的根. 令 $a_0,\cdots,a_m\in\mathbf{F}$, 使得 $a_m\neq 0$ 并且

$$p(z)=a_0+a_1z+a_2z^2+\cdots+a_mz^m,\quad z\in\mathbf{F}.$$

因为 $p(\lambda)=0$, 所以

$$0 = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_m\lambda^m.$$

将上面的两个等式相减, 得

$$p(z) = a_1(z-\lambda) + a_2(z^2-\lambda^2) + \cdots + a_m(z^m-\lambda^m), \quad z \in \mathbf{F}.$$

对每个 $j = 2, \cdots, m$, 都有

$$z^j - \lambda^j = (z-\lambda)q_{j-1}(z), \quad z \in \mathbf{F},$$

其中 q_{j-1} 是某个 $j-1$ 次多项式 (具体地, 取 $q_{j-1}(z) = z^{j-1} + z^{j-2}\lambda + \cdots + z\lambda^{j-2} + \lambda^{j-1}$). 因此

$$p(z) = (z-\lambda)\underbrace{(a_1 + a_2q_1(z) + \cdots + a_mq_{m-1}(z))}_{q(z)}, \quad z \in \mathbf{F}.$$

显然 q 是 $m-1$ 次多项式. ■

现在我们可以证明多项式不会有太多的根.

4.3 推论: 设 $p \in \mathcal{P}(\mathbf{F})$ 是 m 次多项式, $m \geqslant 0$, 则 p 在 $\mathbf{F}$ 中最多有 m 个互不相同的根.

证明: 若 $m = 0$, 则 $p(z) = a_0 \neq 0$, 故 p 没有根. 若 $m = 1$, 则 $p = a_0 + a_1z$, 其中 $a_1 \neq 0$, 故 p 恰好有一个根, 即 $-a_0/a_1$. 现在设 $m > 1$. 对 m 用归纳法, 假设每个 $m-1$ 次多项式最多有 $m-1$ 个不同的根. 如果 p 在 $\mathbf{F}$ 中没有根, 则结论成立. 如果 p 有一个根 $\lambda \in \mathbf{F}$, 则由 4.1, 存在一个 $m-1$ 次多项式 q 使得

$$p(z) = (z-\lambda)q(z), \quad z \in \mathbf{F}.$$

上面的等式表明: 若 $p(z) = 0$, 则 $z = \lambda$ 或者 $q(z) = 0$. 也就是说, p 的根是由 λ 和 q 的根组成的. 由归纳法假设, q 在 $\mathbf{F}$ 中至多有 $m-1$ 个不同的根. 因此 p 在 $\mathbf{F}$ 中至多有 m 个根. ■

下一个结果表明, 如果一个多项式恒等于 0, 那么它的所有系数一定都等于 0.

4.4 推论： 设 $a_0, \cdots, a_m \in \mathbf{F}$. 如果

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m = 0, \quad z \in \mathbf{F},$$

则 $a_0 = \cdots = a_m = 0$.

证明： 假设对所有 $z \in \mathbf{F}$ 都有 $a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$ 等于 0. 根据 4.3, 任何非负整数都不可能是这个多项式的次数. 于是所有的系数都等于 0. ■

上面的推论表明, 对任意非负整数 m, $(1, z, \cdots, z^m)$ 在 $\mathcal{P}(\mathbf{F})$ 中都是线性无关的. 我们早就注意到了这一点 (在第 2 章), 但至此才给出了完整的证明. 这个线性无关性意味着每个多项式都可以唯一地表示成形如 z^j 的函数的线性组合. 特别地, 多项式的次数是唯一的.

如果 p 和 q 都是非负整数, 并且 $p \neq 0$, 则存在非负整数 s 和 r 使得

$$q = sp + r.$$

并且 $r < p$. 考虑 q 除以 p, 则可得商 s 及余数 r. 下面对多项式证明类似的结果.

令 $\deg p$表示多项式 p 的次数. 下一个结果通常称为带余除法, 但这里的表述并不是一个真正的算法, 而只是一个有用的引理.

可以认为, 4.6 给出了 p 除 q 时所得的余项为 r.

4.5 带余除法 (Division Algorithm)： 设 $p, q \in \mathcal{P}(\mathbf{F})$, 并且 $p \neq 0$, 则存在多项式 $s, r \in \mathcal{P}(\mathbf{F})$, 使得

4.6
$$q = sp + r$$

并且 $\deg r < \deg p$.

证明： 取 $s \in \mathcal{P}(\mathbf{F})$ 使得 $q - sp$ 的次数最小. 令 $r = q - sp$. 于是 4.6 成立, 下面只需证明 $\deg r < \deg p$. 假设 $\deg r \geqslant \deg p$. 如果 $c \in \mathbf{F}$ 并且 j 是一个非负整数, 那么

$$q - (s + cz^j)p = r - cz^j p.$$

取 j 和 c 使得上面等式右端多项式的次数小于 $\deg r$ (具体地,

取 $j = \deg r - \deg p$, 然后取 c 使得在 r 和 $cz^j p$ 中 $z^{\deg r}$ 的系数相同). 这与 s 的取法矛盾, 因为我们已经把 s 取成了使得表达式 $q - sp$ 具有最小次数的多项式. ■

§4.2 复 系 数

到目前为止, 通过约定 **F** 代表 **R** 或 **C**, 我们已经同时处理了复系数多项式和实系数多项式. 现在我们将会看到这两种情形之间的区别. 本节讨论复系数多项式. 下一节将利用这些关于复系数多项式的结果来证明实系数多项式的相应结果.

虽然本章没有线性代数内容, 但是到目前为止这些结果都还是用代数证明的. 下一个结果, 虽然称为代数学基本定理, 但是它的证明却需要分析知识. 这里给出的简短证明用到了复分析的工具. 如果你还没有学过复分析方面的课程, 那么这个证明对你而言就不太有意义. 要是这样的话, 只要将代数学基本定理当作一种事实来接受就可以了, 这种事实的证明所用到的高级工具往往在后续的课程中才能学到.

4.7 代数学基本定理 (Fundamental Theorem of Algebra): *每个不是常数的复系数多项式都有根.*

这是一个存在性定理. 二次求根公式明确地给出了 2 次多项式的根. 3 次或 4 次多项式也有类似的但要复杂得多的公式. 5 次以上的多项式就没有这样的公式了.

证明: 设 p 为不是常数的复系数多项式. 假设 p 没有根, 则 $1/p$ 是 **C** 上的解析函数. 进而当 $z \to \infty$ 时, $p(z) \to \infty$, 这说明当 $z \to \infty$ 时, $1/p \to 0$. 因此 $1/p$ 是 **C** 上的有界解析函数. 根据 Liouville 定理, 任何这样的函数都一定是常数. 但若 $1/p$ 是常数, 则 p 是常数, 这与 p 不是常数的假设矛盾. ■

由代数学基本定理可以得到下面的关于复系数多项式分解的结果. 注意到在这个分解中, 数 $\lambda_1, \cdots, \lambda_m$恰好是 p 的所有根, 因为只有当 z 取这些值时才能使 4.9 的右端等于 0.

4.8 推论: *如果 $p \in \mathcal{P}(\mathbf{C})$ 是非常数多项式, 则 p 可以唯一分解 (除因子的次序之外) 成如下形式*

4.9 $$p(z) = c(z - \lambda_1) \cdots (z - \lambda_m),$$

其中 $c, \lambda_1, \cdots, \lambda_m \in \mathbf{C}$.

证明：设 $p \in \mathcal{P}(\mathbf{C})$, m 表示 p 的次数. 对 m 用归纳法. 若 $m = 1$, 则分解显然存在且唯一. 假设 $m > 1$, 并设对于 $m-1$ 次多项式分解存在且唯一.

我们首先证明 p 的分解存在. 由代数学基本定理 (4.7), p 有一个根 λ. 由 4.1, 存在 $m-1$ 次多项式 q 使得

$$p(z) = (z-\lambda)q(z), \quad z \in \mathbf{C}.$$

由归纳法假设, q 有想要的分解, 将这一分解代入上面的等式即可得 p 的分解.

现在考虑唯一性问题. 显然 c 是由 4.9 唯一确定的 —— 它一定等于 p 中 z^m 的系数. 因此我们只需证明除了次序之外, $\lambda_1, \cdots, \lambda_m$ 是唯一确定的. 如果

$$(z-\lambda_1)\cdots(z-\lambda_m) = (z-\tau_1)\cdots(z-\tau_m), \quad z \in \mathbf{C},$$

那么, 因为当 $z = \lambda_1$ 时上面等式的左端等于 0, 所以右端一定有一个 τ 等于 λ_1. 重新排列下标, 可设 $\tau_1 = \lambda_1$. 现在对于 $z \neq \lambda_1$, 上式两端都除以 $z - \lambda_1$, 则

$$(z-\lambda_2)\cdots(z-\lambda_m) = (z-\tau_2)\cdots(z-\tau_m)$$

对除 $z = \lambda_1$ 之外的所有 $z \in \mathbf{C}$ 都成立. 事实上, 上面的等式一定对所有的 $z \in \mathbf{C}$ 都成立; 否则, 从等式的左端减去右端将得到一个具有无限多个根的非零多项式. 由上面的等式及我们的归纳法假设可知, 除了次序之外, 诸 λ 与诸 τ 是相同的, 这就完成唯一性的证明. ■

§4.3 实 系 数

在讨论实系数多项式之前, 我们需要学习更多的复数知识.

设 $z = a + bi$, 其中 a 和 b 都是实数, 则 a 称为 z 的**实部** (**real part**), 记为 $\operatorname{Re} z$, 而 b 称为 z 的**虚部** (**imaginary part**), 记为 $\operatorname{Im} z$. 于是对任意复数 z, 都有

$$z = \text{Re } z + (\text{Im } z)i.$$

复数 z 的**复共轭** (**complex conjugate**), 记为 $\bar{z}$, 定义为

$$\bar{z} = \text{Re } z - (\text{Im } z)i.$$

注意, $z = \bar{z}$ 当且仅当 z 是一个实数.

例如 $\overline{2+3i} = 2 - 3i$.

复数 z 的**绝对值** (**absolute value**), 记为 $|z|$, 定义为

$$|z| = \sqrt{(\text{Re } z)^2 + (\text{Im } z)^2}.$$

例如, $|1+2i| = \sqrt{5}$. 注意, $|z|$ 总是非负的.

你应该验证, 实部、虚部、绝对值和复共轭具有以下性质:

实部的可加性 (additivity of real part)

对所有 $w, z \in \mathbf{C}$ 都有 $\text{Re } (w+z) = \text{Re } w + \text{Re } z$;

虚部的可加性 (additivity of imaginary part)

对所有 $w, z \in \mathbf{C}$ 都有 $\text{Im } (w+z) = \text{Im } w + \text{Im } z$;

z 与 $\bar{z}$ 的和 (sum of z and $\bar{z}$)

对所有 $z \in \mathbf{C}$ 都有 $z + \bar{z} = 2\text{Re } z$;

z 与 $\bar{z}$ 的差 (difference of z and $\bar{z}$)

对所有 $z \in \mathbf{C}$ 都有 $z - \bar{z} = 2(\text{Im } z)i$;

z 与 $\bar{z}$ 的积 (product of z and $\bar{z}$)

对所有 $z \in \mathbf{C}$ 都有 $z\bar{z} = |z|^2$;

复共轭的可加性 (additivity of complex conjugate)

对所有 $w, z \in \mathbf{C}$ 都有 $\overline{w+z} = \bar{w} + \bar{z}$;

复共轭的可乘性 (multiplicativity of complex conjugate)

对所有 $w, z \in \mathbf{C}$ 都有 $\overline{wz} = \bar{w}\bar{z}$;

共轭的共轭 (conjugate of conjugate)

对所有 $z \in \mathbf{C}$ 都有 $\bar{\bar{z}} = z$;

绝对值的可乘性 (multiplicativity of absolute value)

对所有 $w, z \in \mathbf{C}$ 都有 $|wz| = |w||z|$.

在下一个结果中，我们需要把实系数的多项式看成 $\mathcal{P}(\mathbf{C})$ 中的元素. 因为每个实数都是复数，所以这是有意义的.

实系数多项式可能没有实根. 例如，多项式 $1+x^2$ 就没有实根. 在后面几章将会看到，代数学基本定理对 $\mathbf{R}$ 不成立，这说明了实和复向量空间上运算的差别.

4.10 命题： 设 p 是实系数多项式. 如果 $\lambda \in \mathbf{C}$ 是 p 的根，则 $\bar{\lambda}$ 也是 p 的根.

证明： 设

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m,$$

其中 $a_0, \cdots, a_m$ 是实数，并设 $\lambda \in \mathbf{C}$ 是 p 的一个根，则

$$a_0 + a_1\lambda + \cdots + a_m\lambda^m = 0.$$

等式两端取复共轭，得

$$a_0 + a_1\bar{\lambda} + \cdots + a_m\bar{\lambda}^m = 0,$$

此处我们用到了刚刚列出的复共轭的一些基本性质. 上面的等式表明 $\bar{\lambda}$ 是 p 的一个根. ■

我们想要证明实系数多项式的分解定理. 为此，先来描述可以写成两个 1 次实系数多项式之积的 2 次实系数多项式.

考虑一下二次求根公式和此命题的联系.

4.11 命题： 设 $\alpha, \beta \in \mathbf{R}$，则存在形如

4.12 $$x^2 + \alpha x + \beta = (x - \lambda_1)(x - \lambda_2), \quad \lambda_1, \lambda_2 \in \mathbf{R}$$

的多项式分解当且仅当 $\alpha^2 \geqslant 4\beta$.

证明： 注意到

4.13 $$x^2 + \alpha x + \beta = \left(x + \frac{\alpha}{2}\right)^2 + \left(\beta - \frac{\alpha^2}{4}\right).$$

首先假设 $\alpha^2 < 4\beta$，则对每个 $x \in \mathbf{R}$，上式右端显然都是正的，因此多项式 $x^2 + \alpha x + \beta$ 没有实根. 于是不存在形如 4.12 的分解.

反之，设 $\alpha^2 \geqslant 4\beta$，则存在一个实数 c 使得 $c^2 = \frac{\alpha^2}{4} - \beta$. 由

4.13 得

$$x^2+\alpha x+\beta=\left(x+\frac{\alpha}{2}\right)^2-c^2$$

$$=\left(x+\frac{\alpha}{2}+c\right)\left(x+\frac{\alpha}{2}-c\right),$$

这就是我们要的分解. ■

在下面的定理中, 每个形如 $x^2+\alpha_j x+\beta_j$ 的项, 其中 $\alpha_j^2<4\beta_j$, 都不能分解成两个次数为 1 的实系数多项式的乘积(由 4.11). 注意到在下面的分解中, 数 $\lambda_1,\cdots,\lambda_m$ 恰好是 p 的所有实根, 因为只有当 x 取这些实数值时下面等式的右端才等于 0.

4.14 定理: 如果 $p\in\mathcal{P}(\mathbf{R})$ 是非常数多项式, 则 p 可以唯一 (除因子的次序之外) 分解成如下形式

$$p(x)=c(x-\lambda_1)\cdots(x-\lambda_m)(x^2+\alpha_1 x+\beta_1)\cdots(x^2+\alpha_M x+\beta_M),$$

其中 $c,\lambda_1,\cdots,\lambda_m\in\mathbf{R}$, $(\alpha_1,\beta_1),\cdots,(\alpha_M,\beta_M)\in\mathbf{R}^2$, 并且对每个 j 都有 $\alpha_j^2<4\beta_j$.

这里 m 或 M 都有可能等于 0.

证明: 设 $p\in\mathcal{P}(\mathbf{R})$ 是非常数多项式. 我们可以把 p 看作 $\mathcal{P}(\mathbf{C})$ 中的元素 (因为每个实数都是复数). 证明的思路是, 把 p 看作复系数多项式再利用 p 的分解 4.9. 由 4.10 可知, p 的非实数的复根总是成对出现的. 于是, 如果 p 作为 $\mathcal{P}(\mathbf{C})$ 中的元素, 它的分解包含形如 $(x-\lambda)$ 的项, 其中 λ 是非实数的复数, 那么 $(x-\bar{\lambda})$ 也是该分解中的一项. 将这两项相乘就得到了所需形式的二次项.

上一段简要描述的思路基本上给出了分解存在性的证明. 但是, 有一点需要注意, 假设 λ 是一个非实数的复根, 并且 $(x-\lambda)$ 是 $p\in\mathcal{P}(\mathbf{C})$ 的分解中的一项. 根据 4.10, $(x-\bar{\lambda})$ 也是这个分解中的一项, 但是 4.10 并没说这两个因子出现的次数相同, 而这是使上面的思路可行所必需的. 但这不成问题: 存在多项式 $q\in\mathcal{P}(\mathbf{C})$, 使得 q 的次数比 p 的次数小 2, 并且

$$p(x)=(x-\lambda)(x-\bar{\lambda})q(x)$$

$$= \left(x^2 - 2(\text{Re }\lambda)x + |\lambda|^2\right)q(x).$$

如果我们能够证明 q 的系数都是实数, 那么通过对 p 的次数用归纳法就可以断定: $(x-\lambda)$ 在 p 的分解中出现的次数恰好等于 $(x-\bar{\lambda})$ 出现的次数.

此处分母一定不为零, 因为 $x^2-2(\text{Re }\lambda)x+|\lambda|^2$ 的两个根 λ 和 $\bar{\lambda}$ 都不是实数.

要证明 q 是实系数的, 从上面的方程解出 q 可得

$$q(x) = \frac{p(x)}{x^2 - 2(\text{Re }\lambda)x + |\lambda|^2}, \quad x \in \mathbf{R}.$$

由上式可知, 对所有 $x \in \mathbf{R}$ 都有 $q(x) \in \mathbf{R}$. 把 q 写成

$$q(x) = a_0 + a_1 x + \cdots + a_{n-2}x^{n-2},$$

其中 $a_0, \cdots, a_{n-2} \in \mathbf{C}$, 从而对所有 $x \in \mathbf{R}$ 都有

$$0 = \text{Im }q(x) = (\text{Im }a_0) + (\text{Im }a_1)x + \cdots + (\text{Im }a_{n-2})x^{n-2}.$$

由此得 $\text{Im }a_0, \cdots, \text{Im }a_{n-2}$ 都等于 0 (由 4.4). 因此 q 的所有系数都是实数, 于是分解存在.

现在我们来考虑分解的唯一性问题. p 的形如 $x^2+\alpha x+\beta$ 的因子, 其中 $\alpha^2 < 4\beta$, 可以唯一写成 $(x-\lambda)(x-\bar{\lambda})$, 其中 $\lambda \in \mathbf{C}$. 稍加思索便知, 如果 p 作为 $\mathcal{P}(\mathbf{R})$ 中元素有两个不同的分解, 那么 p 作为 $\mathcal{P}(\mathbf{C})$ 中元素也有两个不同的分解, 与 4.8 矛盾. ■

习　题

1. 设 m 和 n 都是正整数, 并且 $m \leqslant n$. 证明存在一个多项式 $p \in \mathcal{P}_n(\mathbf{F})$, 其恰有 m 个不同的根.

2. 设 $z_1, \cdots, z_{m+1}$ 是 $\mathbf{F}$ 中的不同元素, 并且 $w_1, \cdots, w_{m+1} \in \mathbf{F}$. 证明: 存在唯一一个多项式 $p \in \mathcal{P}_m(\mathbf{F})$, 使得对 $j = 1, \cdots, m+1$ 都有
$$p(z_j) = w_j.$$

3. 证明: 如果 $p, q \in \mathcal{P}(\mathbf{F})$, 并且 $p \neq 0$, 那么存在唯一一对多项式 $s, r \in \mathcal{P}(\mathbf{F})$ 使得
$$q = sp + r$$
并且 $\deg r < \deg p$. 也就是说, 可以给带余除法 (4.5) 添加关于唯一性的陈述.

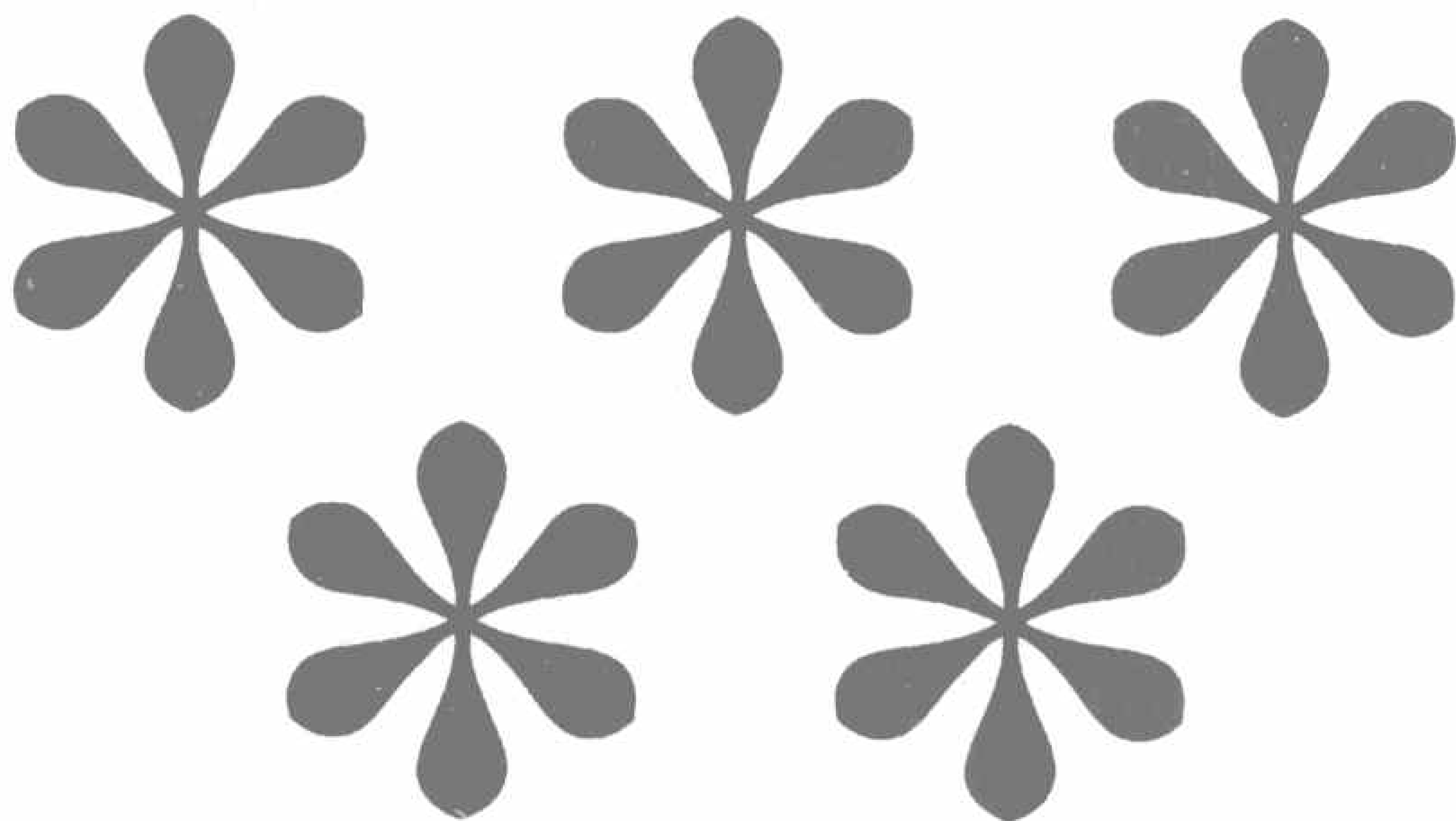
4. 设 $p \in \mathcal{P}(\mathbf{C})$ 的次数为 m. 证明: p 有 m 个互不相同的根当且仅当 p 和它的导数 p' 没有公共根.

5. 证明奇数次的实系数多项式都有实根.

第 5 章　本征值与本征向量

第 3 章研究了从一个向量空间到另一个向量空间的线性映射. 现在开始研究从一个向量空间到其自身的线性映射. 对这种线性映射的研究构成了线性代数最深刻、最重要的部分. 这方面的大部分结果对无限维向量空间都不成立, 因此我们只考虑有限维向量空间. 为了避免平凡性, 我们也不考虑向量空间 $\{\mathbf{0}\}$. 因此, 我们作如下假设:

$\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$.
V 是 $\mathbf{F}$ 上的有限维非零向量空间.



§5.1　不变子空间

本章我们要引进一些工具, 这将有助于我们理解算子的结构. 回想一下, 算子是从一个向量空间到其自身的线性映射, V 上算子的集合记为 $\mathcal{L}(V)$, 即 $\mathcal{L}(V) = \mathcal{L}(V, V)$.

我们来看看如何能更好地理解算子. 设 $T \in \mathcal{L}(V)$. 如果 V 有直和分解

5.1 $$V = U_1 \oplus \cdots \oplus U_m,$$

其中每个 U_j 都是 V 的真子空间, 那么, 为了理解 T 的性质, 我们只需理解每个 $T|_{U_j}$ 的性质, 这里 $T|_{U_j}$表示把 T 限制到更小的定义域 U_j 上. 因为 U_j 是比 V 更小的向量空间, 所以处理 $T|_{U_j}$ 应该比处理 T 更容易. 但是, 如果想要使用算子研究中的一些有效工具 (比如取幂), 那么就会有一个问题: $T|_{U_j}$ 可能不把 U_j 映到自身; 也就是说, $T|_{U_j}$ 可能不是 U_j 上的算子. 因此, 我们只考虑形如 5.1 并且具有如下性质的分解: T 把其中的每个 U_j 都映到自身.

被算子映到自身的子空间十分重要, 应当有个名字. 因而, 对于 $T \in \mathcal{L}(V)$ 和 V 的子空间 U, 如果对每个 $u \in U$ 都有 $Tu \in U$, 则称 U 在 T 下是**不变的 (invariant)**. 也就是说, $T|_U$ 是 U 上的算子当且仅当 U 在 T 下是不变的. 例如, 若 T 是 $\mathcal{P}_7(\mathbf{R})$ 上的微分算子, 则 $\mathcal{P}_4(\mathbf{R})$ (它是 $\mathcal{P}_7(\mathbf{R})$ 的子空间) 在 T 下是不变的, 这是因为次数不超过 4 的多项式的导数仍是次数不超过 4 的多项式.

在泛函分析中, 最著名的尚未解决的问题叫做不变子空间问题. 它研究无限维向量空间上算子的不变子空间.

我们来看不变子空间的一些简单例子. 设 $T \in \mathcal{L}(V)$, 显然, $\{\mathbf{0}\}$ 在 T 下是不变的; 整个空间 V 在 T 下也是不变的. T 是不是一定有不同于 $\{\mathbf{0}\}$ 和 V 的不变子空间呢? 我们以后会看到, 这个问题对于维数大于 1 的复向量空间上的算子和维数大于 2 的实向量空间上的算子都有肯定的答案.

若 $T \in \mathcal{L}(V)$, 则 $\operatorname{null} T$ 在 T 下是不变的 (证明: 若 $\boldsymbol{u} \in \operatorname{null} T$, 则 $T\boldsymbol{u} = \mathbf{0}$, 故 $T\boldsymbol{u} \in \operatorname{null} T$); $\operatorname{range} T$ 在 T 下也是不变的 (证明: 若 $\boldsymbol{u} \in \operatorname{range} T$, 则由值域的定义可知, $T\boldsymbol{u}$ 含于 $\operatorname{range} T$). 虽然 $\operatorname{null} T$ 和 $\operatorname{range} T$ 在 T 下都是不变的, 但是这并未给不同于 $\{\mathbf{0}\}$ 和 V 的不变子空间的存在性问题提供简单答案, 这是因

为, null$T$ 可能等于 $\{\mathbf{0}\}$, 而 range$T$ 可能等于 V (当 T 可逆时, 就是这样).

我们以后还会回过头来更深入地研究不变子空间. 现在先来研究最简单的非平凡不变子空间 —— 1 维不变子空间.

算子在 1 维不变子空间上有什么样的性质呢? V 的 1 维子空间是容易描述的. 任取非零向量 $\boldsymbol{u} \in V$, 并设 U 等于 $\boldsymbol{u}$ 的标量倍之集:

5.2 $$U = \{a\boldsymbol{u} : a \in \mathbf{F}\}.$$

则 U 是 V 的 1 维子空间, 而且 V 的每个 1 维子空间都具有这种形式. 若 $\boldsymbol{u} \in V$, 并且 5.2 所定义的子空间 U 在 $T \in \mathcal{L}(V)$ 下是不变的, 则 $T\boldsymbol{u}$ 必含于 U, 因此必有标量 $\lambda \in \mathbf{F}$ 使得 $T\boldsymbol{u} = \lambda\boldsymbol{u}$. 反之, 若 $\boldsymbol{u}$ 是 V 中的非零向量, 并且有某个 $\lambda \in \mathbf{F}$ 使得 $T\boldsymbol{u} = \lambda\boldsymbol{u}$, 则 5.2 所定义的 V 的 1 维子空间 U 在 T 下是不变的.

这些子空间和 Herbert Marcuse 的名著《**一维人**》的主题没多大联系.

我们刚才已经看到, 方程

5.3 $$T\boldsymbol{u} = \lambda\boldsymbol{u},$$

和 1 维不变子空间密切相关, 而且十分重要, 因此我们给满足此方程的向量 $\boldsymbol{u}$ 和标量 λ 分别起一个特殊的名字. 具体地, 对于 $T \in \mathcal{L}(V)$ 和标量 $\lambda \in \mathbf{F}$, 如果有非零向量 $\boldsymbol{u} \in V$ 使得 $T\boldsymbol{u} = \lambda\boldsymbol{u}$, 则称 λ 为 T 的**本征值 (eigenvalue)**. 我们必须要求 $\boldsymbol{u}$ 是非零的, 这是因为, 对于 $\boldsymbol{u} = \mathbf{0}$, 每个标量 $\lambda \in \mathbf{F}$ 都满足 5.3. 上面这些解释表明, T 有 1 维不变子空间当且仅当 T 有本征值.

eigenvalue 是个奇怪的词, 它一半是德文, 一半是英文. 德文形容词 eigen 的意思是"特有的". 有些数学家使用术语**特征值 (characteristic value)** 而不使用**本征值 (eigenvalue)**.

方程 $T\boldsymbol{u} = \lambda\boldsymbol{u}$ 等价于 $(T - \lambda I)\boldsymbol{u} = \mathbf{0}$, 故 λ 是 T 的本征值当且仅当 $T - \lambda I$ 不是单的. 根据 3.21, λ 是 T 的本征值当且仅当 $T - \lambda I$ 不可逆, 当且仅当 $T - \lambda I$ 不是满的.

设 $T \in \mathcal{L}(V)$, 并且 $\lambda \in \mathbf{F}$ 是 T 的本征值. 如果向量 $\boldsymbol{u} \in V$ 满足 $T\boldsymbol{u} = \lambda\boldsymbol{u}$, 则称 $\boldsymbol{u}$ 是 T (相应于 λ) 的**本征向量 (eigenvector)**. 因为 5.3 等价于 $(T - \lambda I)\boldsymbol{u} = \mathbf{0}$, 所以 T 的相应于 λ 的本征向量之集等于 null$(T - \lambda I)$. 特别地, T 的相应于 λ 的本征向量之集是 V 的子空间.

有些教材并不把 **0** 当成本征向量. 按照本书采用的定义, 相应于一个固定本征值的本征向量之集是子空间.

我们来看一些本征值和本征向量的例子. 若 $a \in \mathbf{F}$, 则 aI 仅有一个本征值, 即 a, 并且对于此本征值, 每个向量都是本征向量.

再看一个更复杂的例子, 考虑如下定义的算子 $T \in \mathcal{L}(\mathbf{F}^2)$:

5.4 $$T(w, z) = (-z, w).$$

若 $\mathbf{F} = \mathbf{R}$, 则此算子有很好的几何解释: T 是 $\mathbf{R}^2$ 中绕原点的逆时针 $90°$ 旋转. 一个算子有本征值当且仅当在定义域中存在非零向量能被该算子映成此向量的标量倍. $\mathbf{R}^2$ 中非零向量的旋转显然不能等于此向量的标量倍. 结论: 若 $\mathbf{F} = \mathbf{R}$, 则由 5.4 定义的算子 T 没有本征值. 但是, 在 $\mathbf{F} = \mathbf{C}$ 时情况就不一样了. 为了找出 T 的本征值, 我们必须找出使得方程

$$T(w, z) = \lambda(w, z)$$

有不同于 $w = z = 0$ 的解的标量 λ. 对于由 5.4 定义的 T, 上面的方程等价的于联立方程

5.5 $$-z = \lambda w, \quad w = \lambda z.$$

把第二个方程代入第一个方程可得

$$-z = \lambda^2 z.$$

现在 z 不能等于 0 (否则, 由 5.5 可知 $w = 0$, 而我们要找的 5.5 的解应使 (w, z) 不是 0 向量), 故由上面的方程可得

$$-1 = \lambda^2.$$

这个方程的解是 $\lambda = i$ 和 $\lambda = -i$. 你应该能轻松地验证, i 和 $-i$ 都是 T 的本征值. 的确, 相应于本征值 i 的本征向量是形如 $(w, -wi)$ 的向量, 而相应于本征值 $-i$ 的本征向量是形如 (w, wi) 的向量, 其中 $w \in C$.

现在我们来证明: 相应于不同本征值的非零本征向量是线性无关的.

5.6 定理: *设 $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 是 T 的互不相同的本征值, $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m$ 是相应的非零本征向量, 则 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 线性无关.*

证明: 假设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 线性相关. 设 k 是使得下式成立的最小正整数:

5.7
$$\boldsymbol{v}_k \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1});$$

由线性相关性引理 (2.4) 可知, 具有这种性质的 k 一定存在. 于是, 有 $a_1, \cdots, a_{k-1} \in \mathbf{F}$ 使得

5.8
$$\boldsymbol{v}_k = a_1\boldsymbol{v}_1 + \cdots + a_{k-1}\boldsymbol{v}_{k-1}.$$

把 T 作用到这个等式的两端可得

$$\lambda_k\boldsymbol{v}_k = a_1\lambda_1\boldsymbol{v}_1 + \cdots + a_{k-1}\lambda_{k-1}\boldsymbol{v}_{k-1}.$$

在 5.8 的两端乘以 λ_k, 然后减去上式, 得

$$0 = a_1(\lambda_k - \lambda_1)\boldsymbol{v}_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1})\boldsymbol{v}_{k-1}.$$

因为我们选取的 k 是满足 5.7 的最小正整数, 所以 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1})$ 线性无关. 于是, 由上面的等式可知, 这些 a 都是 0 (回忆一下, λ_k 不等于 $\lambda_1, \cdots, \lambda_{k-1}$ 中的任何一个). 但是, 这意味着 $\boldsymbol{v}_k$ 等于$\mathbf{0}$ (参见 5.8), 与这些 $\boldsymbol{v}$ 都不为零的假设矛盾. 所以 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 线性相关的假设不成立. ■

下面的推论表明, 算子的互异本征值的个数不会多于向量空间的维数.

5.9 推论: *V 上的每个算子最多有 $\dim V$ 个互不相同的本征值.*

证明: 设 $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 是 T 的互不相同的本征值, $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m$ 是相应的非零本征向量. 定理 5.6 表明, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m)$ 线性无关. 因此 $m \leqslant \dim V$ (参见 2.6). ■

§5.2 多项式对算子的作用

算子 (它把一个向量空间映到自身) 理论要比线性映射的理论丰富, 主要原因是算子能自乘为幂. 本节, 我们定义多项式对算子作用的记号和主要概念.

若 $T \in \mathcal{L}(V)$, 则 TT 有意义, 并且也含于 $\mathcal{L}(V)$. 通常用 T^2 来代替 TT. 更一般地, 若 m 是正整数, 则 T^m 定义如下:

$$T^m = \underbrace{T \cdots T}_{m\text{个}}.$$

为了方便, 我们把 T^0 定义为 V 上的恒等算子 I.

回想一下第 3 章, 若 T 是可逆算子, 则 T 的逆记为 T^{-1}. 若 m 是正整数, 则把 T^{-m} 定义为 $(T^{-1})^m$.

你应该验证, 若 T 是算子, 则

$$T^m T^n = T^{m+n}, \quad (T^m)^n = T^{mn},$$

其中 m 和 n 当 T 可逆时是任意整数, 而当 T 不可逆时是非负整数.

如果 $T \in \mathcal{L}(V)$, 并且 $p \in \mathcal{P}(\mathbf{F})$ 是如下多项式

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m, \quad z \in \mathbf{F},$$

则 $p(T)$ 是如下定义的算子

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

例如, 若 p 是多项式, $p(z) = z^2$, $z \in \mathbf{F}$, 则 $p(T) = T^2$. 这是符号 p 的新用法, 因为我们把它作用于算子, 而不仅仅作用于 $\mathbf{F}$ 的元素. 对于一个固定的算子 $T \in \mathcal{L}(V)$, 你应该验证, 由 $p \mapsto p(T)$ 所给出的从 $\mathcal{P}(\mathbf{F})$ 到 $\mathcal{L}(V)$ 的函数是线性的.

若 p 和 q 都是多项式, 它们的系数都在 $\mathbf{F}$ 中, 则 pq 是如下定义的多项式

$$(pq)(z) = p(z)q(z), \quad z \in \mathbf{F}.$$

你应该验证下面这个漂亮的乘法性质: 若 $T \in \mathcal{L}(V)$, 则对系数含于 $\mathbf{F}$ 的任意多项式 p 和 q 都有

$$(pq)(T) = p(T)q(T).$$

注意, T 的任意两个多项式都可交换, 即 $p(T)q(T) = q(T)p(T)$, 这是因为

$$p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T).$$

§5.3 上三角矩阵

我们现在给出复向量空间上算子的中心结果之一.

5.10 定理: *有限维非零复向量空间上的每个算子都有本征值.*

将此定理的这个简单证明与使用行列式的标准证明比较一下. 标准证明必须首先定义行列式这个难懂的概念, 然后必须证明行列式为 0 的算子不可逆, 还需要定义特征多项式, 到这时才能着手这个定理的证明: 根本无法从中领会为什么定理会成立.

证明: 设 V 是 n 维复向量空间, $n > 0$, 并设 $T \in \mathcal{L}(V)$. 取 $\boldsymbol{v} \in V$ 使得 $\boldsymbol{v} \neq \mathbf{0}$. 因为 V 是 n 维的, 所以 $n+1$ 个向量

$$(\boldsymbol{v}, T\boldsymbol{v}, T^2\boldsymbol{v}, \cdots, T^n\boldsymbol{v})$$

不可能是线性无关的. 于是, 有不全为 0 的复数 $a_0, \cdots, a_n$ 使得

$$\mathbf{0} = a_0\boldsymbol{v} + a_1T\boldsymbol{v} + \cdots + a_nT^n\boldsymbol{v}.$$

设 m 是使得 $a_m \neq 0$ 的最大下标. 因为 $\boldsymbol{v} \neq \mathbf{0}$, 所以系数 $a_1, \cdots, a_n$ 不可能都是 0, 故 $0 < m \leqslant n$. 以这些 a 为系数作一个多项式, 并将此多项式分解成 (参见 4.8)

$$a_0 + a_1z + \cdots + a_nz^n = c(z - \lambda_1)\cdots(z - \lambda_m),$$

其中 c 是非零复数, $\lambda_j \in \mathbf{C}$, 并且等式对所有 $z \in \mathbf{C}$ 都成立. 我们有

$$\begin{aligned}\mathbf{0} &= a_0\boldsymbol{v} + a_1T\boldsymbol{v} + \cdots + a_nT^n\boldsymbol{v} \\ &= (a_0I + a_1T + \cdots + a_nT^n)\boldsymbol{v} \\ &= c(T - \lambda_1I)\cdots(T - \lambda_mI)\boldsymbol{v},\end{aligned}$$

这意味着至少有一个 j 使得 $T - \lambda_jI$ 不是单的. 也就是说, T 有本征值. ■

回想一下, 在第 3 章我们讨论了从一个向量空间到另一个向量空间的线性映射的矩阵. 这个矩阵依赖于这两个向量空间基的选取. 既然我们要研究从一个向量空间到其自身的算子, 那么有一个基就够了. 另外, 此时的矩阵也将是方阵, 而非以前所考虑的更一般的长方形矩阵. 具体地, 设 $T \in \mathcal{L}(V)$, 且 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, 则对每个 $k = 1, \cdots, n$ 都有

$$T\boldsymbol{v}_k = a_{1,k}\boldsymbol{v}_1 + \cdots + a_{n,k}\boldsymbol{v}_n,$$

把 $T\boldsymbol{v}_k$ 写成这些 $\boldsymbol{v}$ 的线性组合, 所用的系数就构成了这个矩阵的第 k 列.

其中 $a_{j,k} \in \mathbf{F}$, $j = 1, \cdots, n$. 下面的 $n \times n$ 矩阵

5.11
$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix}$$

称为 T 关于基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 的**矩阵 (matrix)**, 记为 $\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))$; 如果基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 在上下文中是明显的 (例如, 只见到一个基), 则简记为 $\mathcal{M}(T)$.

如果 T 是 $\mathbf{F}^n$ 上的算子, 且未指明基, 则假设所考虑的基是标准基 (其中第 j 个基向量的第 j 个位置是 1, 其他位置都是 0). 可以把 $\mathcal{M}(T)$ 的第 j 列看成是 T 作用于第 j 个基向量.

线性代数的一个中心目标就是要证明: 给定一个算子 $T \in \mathcal{L}(V)$, V 有一个基使得 T 关于此基有比较简单的矩阵. 这种表述比较含糊 ("比较简单" 不是精确的语言), 说得具体一点, 就是让 $\mathcal{M}(T)$ 有很多 0.

如果 V 是复向量空间, 那么我们已经知道, V 有一个基使得 T 关于此基的矩阵的第一列除第一个元素之外全是 0, 这个矩阵形如

我们经常用 $*$ 来表示矩阵的未知元素或与所讨论的问题无关的元素.

$$\begin{bmatrix} \lambda & & \\ 0 & * & \\ \vdots & & \\ 0 & & \end{bmatrix};$$

这里 $*$ 表示除第一列之外的所有其他元素. 为了证明这个事实, 设 λ 是 T 的本征值 (由 5.10 可知, 一定存在), 并且 $\boldsymbol{v}$ 是相应的

非零本征向量. 把 $(\boldsymbol{v})$ 扩充成 V 的基, 则 T 关于此基的矩阵就有上述形式. 马上就会看到, 我们可以选取 V 的一个基使得 T 关于此基的矩阵有更多的 0.

方阵的**对角线 (diagonal)** 由位于从左上角到右下角的直线上的元素组成. 例如, 矩阵 5.11 的对角线由元素 $a_{1,1}, a_{2,2}, \cdots, a_{n,n}$ 组成.

一个矩阵称为**上三角的 (upper triangular)**, 如果位于对角线下方的元素全为 0. 例如, 4×4 矩阵

$$\begin{bmatrix} 6 & 2 & 7 & 5 \\ 0 & 6 & 1 & 3 \\ 0 & 0 & 7 & 9 \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

是上三角的. 上三角矩阵具有如下的典型形式

$$\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix};$$

上面矩阵中的那个 0 表示在这个 $n \times n$ 矩阵中, 位于对角线下方的元素全等于 0. 可以认为上三角矩阵是比较简单的 —— 因为当 n 比较大时, $n \times n$ 上三角矩阵有几乎一半的元素等于 0.

下面的命题所揭示的上三角矩阵与不变子空间之间的这种联系是很有用的.

5.12 命题: 设 $T \in \mathcal{L}(V)$, 并且 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, 则下列等价:

(a) T 关于基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 的矩阵是上三角的;

(b) $T\boldsymbol{v}_k \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k)$, $k = 1, \cdots, n$;

(c) $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k)$ 在 T 下是不变的, $k = 1, \cdots, n$.

证明: 根据定义, 稍加思考即可得 (a) 和 (b) 的等价性. 显然 (c) 蕴含 (b). 因此, 为了完成证明, 只需证明 (b) 蕴含 (c). 假设 (b) 成立. 取定 $k \in \{1, \cdots, n\}$. 由 (b) 可知,

$$T\boldsymbol{v}_1 \in \operatorname{span}(\boldsymbol{v}_1) \subset \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k);$$

$$T\boldsymbol{v}_2 \in \operatorname{span}(\boldsymbol{v}_1, \boldsymbol{v}_2) \subset \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k);$$

$$\vdots$$

$$T\boldsymbol{v}_k \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k).$$

因此, 若 $\boldsymbol{v}$ 是 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k)$ 的线性组合, 则

$$T\boldsymbol{v} \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k).$$

也就是说, $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k)$ 在 T 下是不变的. ■

现在我们可以证明: 对于复向量空间上的每个算子, 向量空间都有一个基使得算子关于此基的矩阵的对角线下方只有 0. 在第 8 章我们将改进这个结果.

此定理在实向量空间上不成立, 这是因为, 若算子关于一个基有上三角矩阵, 则此基中的第一个向量必须是该算子的本征向量. 因此, 若实向量空间上的一个算子没有本征值 (我们已经见过 $\mathbf{R}^2$ 上的一个例子), 则该算子关于任何基都不会有上三角矩阵.

5.13 定理: 设 V 是复向量空间, 并设 $T \in \mathcal{L}(V)$, 则 T 关于 V 的某个基具有上三角矩阵.

证明: 对 V 的维数用归纳法. 显然, 若 $\dim V = 1$, 则结论成立.

现在假设 $\dim V > 1$, 并设对于所有维数比 V 小的复向量空间, 结果都成立. 设 λ 是 T 的任意本征值 (5.10 保证了 T 有本征值), 再设

$$U = \operatorname{range}(T - \lambda I).$$

因为 $T - \lambda I$ 不是满的 (参见 3.21), 所以 $\dim U < \dim V$. 进一步, U 在 T 下是不变的. 为了证明这点, 设 $\boldsymbol{u} \in U$, 则

$$T\boldsymbol{u} = (T - \lambda I)\boldsymbol{u} + \lambda \boldsymbol{u}.$$

显然, $(T - \lambda I)\boldsymbol{u} \in U$ (根据 U 的定义), 并且 $\lambda \boldsymbol{u} \in U$. 因此, 上面的等式表明 $T\boldsymbol{u} \in U$, 故 U 在 T 下是不变的.

因此 $T|_U$ 是 U 上的算子. 由归纳法假设, U 有基 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$ 使得 $T|_U$ 关于此基有上三角矩阵. 因此, 对每个 j 都有 (利用 5.12)

5.14 $$T\boldsymbol{u}_j = (T|_U)(\boldsymbol{u}_j) \in \operatorname{span}(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j).$$

把 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$ 扩充成 V 的基 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$, 则对

每个 k 都有,

$$T\boldsymbol{v}_k = (T - \lambda I)\boldsymbol{v}_k + \lambda\boldsymbol{v}_k.$$

U 的定义表明, $(T - \lambda I)\boldsymbol{v}_k \in U = \operatorname{span}(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m)$. 因此, 由上式得

5.15 $$T\boldsymbol{v}_k \in \operatorname{span}(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k).$$

由 5.14 和 5.15 可得 (利用 5.12), T 关于基 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 有上三角矩阵. ■

如何通过考虑一个算子的矩阵来确定该算子是否可逆呢? 如果我们很幸运地有一个基使得算子关于此基的矩阵是上三角的, 那么问题就会变得很简单, 正如下面的命题所示.

5.16 命题: *假设 $T \in \mathcal{L}(V)$ 关于 V 的某个基有上三角矩阵, 则 T 可逆当且仅当这个上三角矩阵对角线上的元素都不是 0.*

证明: 设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基使得 T 关于此基具有上三角矩阵

5.17 $$\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)) = \begin{bmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}.$$

我们需要证明: T 不可逆当且仅当某个 λ_k 等于 0.

首先证明: 若某个 λ_k 等于 0, 则 T 不可逆. 若 $\lambda_1 = 0$, 则 $T\boldsymbol{v}_1 = \mathbf{0}$ (由 5.17), 故 T 不可逆. 假设 $1 < k \leqslant n$, 并且 $\lambda_k = 0$, 则由 5.17, T 把向量 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1}$ 都映到 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1})$ 内. 因为 $\lambda_k = 0$, 所以矩阵 5.17 还意味着 $T\boldsymbol{v}_k \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1})$. 于是, 我们可以定义一个线性映射 S,

$$S : \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \to \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1})$$

使得 $S\boldsymbol{v} = T\boldsymbol{v}$, $\boldsymbol{v} \in \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k)$. 即 S 是 T 在 $\operatorname{span}(\boldsymbol{v}_1, \cdots,$

$\boldsymbol{v}_k$) 上的限制.

注意到 $\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k)$ 的维数等于 k, 而 $\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1})$ 的维数等于 $k-1$ (因为 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 线性无关). 因为 $\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k)$ 的维数比 $\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1})$ 的维数大, 所以从 $\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k)$ 到 $\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1})$ 的线性映射都不是单的 (参见 3.5). 于是有非零向量 $\boldsymbol{v}\in\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k)$ 使得 $S\boldsymbol{v}=\mathbf{0}$, 从而 $T\boldsymbol{v}=\mathbf{0}$, 故 T 不可逆.

为了证明另一个方面, 假设 T 不可逆. 因而 T 不是单的 (参见 3.21), 并且有非零向量 $\boldsymbol{v}\in V$ 使得 $T\boldsymbol{v}=\mathbf{0}$. 因为 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的基, 所以有

$$\boldsymbol{v}=a_1\boldsymbol{v}_1+\cdots+a_k\boldsymbol{v}_k,$$

其中 $a_1,\cdots,a_k\in\mathbf{F}$, $a_k\neq 0$ (把 $\boldsymbol{v}$ 表示成 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 线性组合, 然后取 k 是系数不为 0 的最大下标). 于是

$$\mathbf{0}=T\boldsymbol{v}$$

$$\begin{aligned}\mathbf{0}&=T(a_1\boldsymbol{v}_1+\cdots+a_k\boldsymbol{v}_k)\\&=(a_1T\boldsymbol{v}_1+\cdots+a_{k-1}T\boldsymbol{v}_{k-1})+a_kT\boldsymbol{v}_k.\end{aligned}$$

最后那个括号中的向量含于 $\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1})$ (由于 5.17 是上三角的). 因此, 由上面的等式得 $a_kT\boldsymbol{v}_k\in\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1})$. 因为 $a_k\neq 0$, 所以乘以 $1/a_k$ 可得, $T\boldsymbol{v}_k\in\text{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1})$. 于是, 若把 $T\boldsymbol{v}_k$ 写成基 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 的线性组合, 则 $\boldsymbol{v}_k$ 的系数是 0, 即 5.17 中的 λ_k 必为 0. ■

通过算子的矩阵来计算其本征值的有效数值技术是有的.

不幸的是, 现在还无法通过一般算子 (关于任意基) 的矩阵来严格计算算子的本征值. 但是, 如果我们很幸运地找到了一个基使得算子关于此基的矩阵是上三角的, 则本征值的计算问题就变得平凡了, 正如下面的命题所示.

5.18 命题: 设 $T\in\mathcal{L}(V)$ 关于 V 的某个基有上三角矩阵, 则这个上三角矩阵对角线上的元素恰好是 T 的所有本征值.

证明： 设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基，并且 T 关于此基有上三角矩阵

$$\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)) = \begin{bmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}.$$

设 $\lambda \in \mathbf{F}$，则

$$\mathcal{M}(T - \lambda I, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)) = \begin{bmatrix} \lambda_1 - \lambda & & & * \\ & \lambda_2 - \lambda & & \\ & & \ddots & \\ 0 & & & \lambda_n - \lambda \end{bmatrix}.$$

因此，$T - \lambda I$ 不可逆当且仅当 λ 等于某个 λ_j (参见 5.16)。也就是说，λ 是 T 的本征值当且仅当 λ 等于某个 λ_j. ■

§5.4 对角矩阵

对角矩阵 (diagonal matrix) 是对角线以外的元素全是 0 的方阵. 例如，

$$\begin{bmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

是对角矩阵. 显然，每个对角矩阵都是上三角的，但是对角矩阵一般要比上三角矩阵有更多的 0.

由算子关于一个基的矩阵的定义立即可得，算子 $T \in \mathcal{L}(V)$ 关于 V 的基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 有对角矩阵

$$\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

当且仅当

$$
\begin{aligned}
T\boldsymbol{v}_1 &= \lambda_1\boldsymbol{v}_1\\
&\vdots\\
T\boldsymbol{v}_n &= \lambda_n\boldsymbol{v}_n.
\end{aligned}
$$

因此, 算子 $T \in \mathcal{L}(V)$ 关于 V 的某个基有对角矩阵当且仅当 V 有一个由 T 的本征向量组成的基.

由 5.18 可知 (或者对于对角矩阵直接给出一个简单证明), 若算子关于某个基有对角矩阵, 则对角线上的元素恰好是该算子的所有本征值.

可惜并非每个算子都关于某个基有对角矩阵. 这种糟糕的情况甚至可以出现在复向量空间上. 例如, 考虑 $T \in \mathcal{L}(\mathbf{C}^2)$, 其定义如下

5.19 $$T(w,z) = (z,0).$$

你应该验证, 0 是这个算子唯一的本征值, 并且相应的本征向量的集合是 1 维子空间 $\{(w,0) \in \mathbf{C}^2 : w \in \mathbf{C}\}$. 于是, T 的线性无关的本征向量不足以组成 2 维空间 $\mathbf{C}^2$ 的一个基. 因此 T 关于 $\mathbf{C}^2$ 的任何基都不会有对角矩阵.

下一个命题表明, 如果一个算子的互异本征值的个数与定义域的维数相同, 则此算子关于某个基有对角矩阵; 但是, 某些本征值较少的算子也可能有对角矩阵 (也就是说, 下一个命题的逆命题不成立). 例如, 3 维空间 $\mathbf{F}^3$ 上, 由

$$T(z_1, z_2, z_3) = (4z_1, 4z_2, 5z_3)$$

定义的算子 T 只有两个本征值 (4 和 5), 但这个算子关于标准基有对角矩阵.

以后会找到使得算子关于某个基有对角矩阵的其他条件 (参见 7.9 和 7.13).

5.20 命题: *若 $T \in \mathcal{L}(V)$ 有 $\dim V$ 个互不相同的本征值, 则 T 关于 V 的某个基有对角矩阵.*

证明: 设 $T \in \mathcal{L}(V)$ 有 $\dim V$ 个互不相同的本征值 $\lambda_1, \cdots, \lambda_{\dim V}$. 对每个 j, 设 $\boldsymbol{v}_j \in V$ 是相应于本征值 λ_j 的非零本征向量. 因为相应于互异本征值的非零本征向量线性无关 (参见

5.6), 所以 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{\dim V})$ 线性无关. V 中由 $\dim V$ 个向量组成的线性无关组是 V 的基 (参见 2.17); 故 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{\dim V})$ 是 V 的基. T 关于这个由本征向量组成的基有对角矩阵. ■

这一节的最后一个命题给出了算子关于某个基有对角矩阵的一些等价条件.

5.21 命题: 设 $T \in \mathcal{L}(V)$, 并设 $\lambda_1, \cdots, \lambda_m$ 是 T 的所有互不相同的本征值, 则下列等价:

(a) T 关于 V 的某个基有对角矩阵;

(b) V 有一个由 T 的本征向量组成的基;

(c) V 有在 T 下不变的 1 维子空间 $U_1, \cdots, U_n$, 使得

$$V = U_1 \oplus \cdots \oplus U_n;$$

(d) $V = \operatorname{null}(T - \lambda_1 I) \oplus \cdots \oplus \operatorname{null}(T - \lambda_m I)$;

(e) $\dim V = \dim \operatorname{null}(T - \lambda_1 I) + \cdots + \dim \operatorname{null}(T - \lambda_m I)$.

对于复向量空间, 我们以后将扩展这个等价性列表 (参见第 8 章习题 16 和习题 23).

证明: 我们已经证明了 (a) 和 (b) 是等价的.

假设 (b) 成立, 则 V 有一个由 T 的本征向量组成的基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$. 对每个 j, 设 $U_j = \operatorname{span}(\boldsymbol{v}_j)$. 显然, 每个 U_j 都是 V 的 1 维子空间, 并且在 T 下是不变的 (因为每个 $\boldsymbol{v}_j$ 都是 T 的本征向量). 因为 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, 所以 V 中每个向量都可以唯一地写成 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 的线性组合. 也就是说, V 中每个向量都可以唯一地写成一个和 $\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_n$, 其中每个 $\boldsymbol{u}_j \in U_j$. 于是, $V = U_1 \oplus \cdots \oplus U_n$. 因此, (b) 蕴含 (c).

现在假设 (c) 成立, 则 V 有在 T 下不变的 1 维子空间 $U_1, \cdots, U_n$, 使得

$$V = U_1 \oplus \cdots \oplus U_n.$$

对每个 j, 设 $\boldsymbol{v}_j$ 是 U_j 中的一个非零向量, 则每个 $\boldsymbol{v}_j$ 都是 T 的本征向量. 因为 V 中每个向量都可以唯一地写成一个和 $\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_n$, 其中每个 $\boldsymbol{u}_j \in U_j$ (故每个 $\boldsymbol{u}_j$ 都是 $\boldsymbol{v}_j$ 的标量倍), 所以 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基. 因此 (c) 蕴含 (b).

现在已经证明了 (a), (b), (c) 都等价. 为了完成证明, 我们将证明 (b) 蕴含 (d), (d) 蕴含 (e), 并且 (e) 蕴含 (b).

假设 (b) 成立, 则 V 有一个由 T 的本征向量组成的基. 于是, V 中每个向量都是 T 的本征向量的线性组合. 因此

5.22 $$V = \mathrm{null}(T-\lambda_1 I) + \cdots + \mathrm{null}(T-\lambda_m I).$$

为了证明上面的和是直和, 设

$$\mathbf{0} = \boldsymbol{u}_1 + \cdots + \boldsymbol{u}_m,$$

其中每个 $\boldsymbol{u}_j \in \mathrm{null}(T-\lambda_j I)$. 因为相应于互异本征值的非零本征向量线性无关, 所以每个 $\boldsymbol{u}_j$ 都等于$\mathbf{0}$ (对上式右端的非零向量之和用 5.6), 从而 5.22 中的和是直和 (利用 1.8). 这就证明了 (b) 蕴含 (d).

由第 2 章习题 17 立即可知 (d) 蕴含 (e).

最后, 假设 (e) 成立, 则

5.23 $$\dim V = \dim\mathrm{null}(T-\lambda_1 I) + \cdots + \dim\mathrm{null}(T-\lambda_m I).$$

在每个 $\mathrm{null}(T-\lambda_j I)$ 中取一个基, 把这些基合在一起, 得到 T 的一组本征向量 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$, 其中 $n = \dim V$ (由 5.23). 为了证明这组向量线性无关, 假设

$$a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n = \mathbf{0},$$

其中 $a_1, \cdots, a_n \in \mathbf{F}$. 对每个 $j = 1, \cdots, m$, 设 $\boldsymbol{u}_j$ 表示使得 $\boldsymbol{v}_k \in \mathrm{null}(T-\lambda_j I)$ 的那些项 $a_k\boldsymbol{v}_k$ 的和, 则每个 $\boldsymbol{u}_j$ 都是 T 的相应于本征值 λ_j 的本征向量, 并且

$$\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_m = \mathbf{0}.$$

因为相应于互异本征值的非零本征向量线性无关, 所以每个 $\boldsymbol{u}_j$ 都等于$\mathbf{0}$ (对上式左端的非零向量之和用 5.6). 由于每个 $\boldsymbol{u}_j$ 都是一些 $a_k\boldsymbol{v}_k$ 的和, 其中的这些 $\boldsymbol{v}_k$ 组成了 $\mathrm{null}(T-\lambda_j I)$ 的基, 从而所有的 a_k 都等于 0. 于是, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 线性无关, 因此是 V 的基 (由 2.17). 这就证明了 (e) 蕴含 (b). ■

§5.5 实向量空间的不变子空间

我们知道, 在复向量空间中, 每个算子都有本征值 (准确陈述参见 5.10), 而且有例子表明类似的结论对于实向量空间不成立. 也就是说, 非零实向量空间上的算子可能没有 1 维的不变子空间, 但是我们现在要证明, 总有 1 维或 2 维的不变子空间.

5.24 定理: *在有限维非零实向量空间中, 每个算子都有 1 维或 2 维的不变子空间.*

证明: 设 V 是 n 维实向量空间, $n > 0$, 并设 $T \in \mathcal{L}(V)$. 取 $\boldsymbol{v} \in V$, 且 $\boldsymbol{v} \neq \mathbf{0}$. 因为 V 是 n 维的, 所以 $n+1$ 个向量

$$(\boldsymbol{v}, T\boldsymbol{v}, T^2\boldsymbol{v}, \cdots, T^n\boldsymbol{v})$$

不可能是线性无关的. 于是, 有不全为 0 的实数 $a_0, \cdots, a_n$ 使得

$$\mathbf{0} = a_0\boldsymbol{v} + a_1T\boldsymbol{v} + \cdots + a_nT^n\boldsymbol{v}.$$

以这些 a 为系数作一个多项式, 并将此多项式分解成 (参见 4.14)

$$\begin{aligned} & a_0 + a_1x + \cdots + a_nx^n \\ = & c(x-\lambda_1)\cdots(x-\lambda_m)(x^2+\alpha_1x+\beta_1)\cdots(x^2+\alpha_Mx+\beta_M), \end{aligned}$$

这里 m 或 M 都有可能等于 0.

其中 c 是非零实数, 每个 λ_j, α_j, β_j 都是实的, $m+M \geqslant 1$, 并且上式对所有 $x \in \mathbf{R}$ 都成立. 因此, 我们有

$$\begin{aligned} \mathbf{0} &= a_0\boldsymbol{v} + a_1T\boldsymbol{v} + \cdots + a_nT^n\boldsymbol{v} \\ &= (a_0I + a_1T + \cdots + a_nT^n)\boldsymbol{v} \\ &= c(T-\lambda_1I)\cdots(T-\lambda_mI)(T^2+\alpha_1T+\beta_1I)\cdots \\ &\quad (T^2+\alpha_MT+\beta_MI)\boldsymbol{v}, \end{aligned}$$

这意味着至少有一个 j 使得 $T-\lambda_jI$ 不是单的或者 $(T^2+\alpha_jT+\beta_jI)$ 不是单的. 若有某个 j 使得 $T-\lambda_jI$ 不是单的, 则 T 有本征值, 因此有 1 维不变子空间. 现在考虑另一种可能. 也就是说, 假设有某个 j 使得 $(T^2+\alpha_jT+\beta_jI)$ 不是单的, 则有非零向量 $\boldsymbol{u} \in V$ 使得

5.25 $$T^2\boldsymbol{u}+\alpha_j T\boldsymbol{u}+\beta_j\boldsymbol{u}=\boldsymbol{0}.$$

为了完成证明, 往证 $\operatorname{span}(\boldsymbol{u},T\boldsymbol{u})$ 在 T 下是不变的, 它显然是 1 维或 2 维的. 为此, 考虑 $\operatorname{span}(\boldsymbol{u},T\boldsymbol{u})$ 中的一个典型元素 $a\boldsymbol{u}+bT\boldsymbol{u}$, 其中 $a,b\in\mathbf{R}$, 则由 5.25 可得

$$\begin{aligned}T(a\boldsymbol{u}+bT\boldsymbol{u})&=aT\boldsymbol{u}+bT^2\boldsymbol{u}\\&=aT\boldsymbol{u}-b\alpha_jT\boldsymbol{u}-b\beta_j\boldsymbol{u}.\end{aligned}$$

上式表明 $T(a\boldsymbol{u}+bT\boldsymbol{u})\in\operatorname{span}(\boldsymbol{u},T\boldsymbol{u})$. 于是, $\operatorname{span}(\boldsymbol{u},T\boldsymbol{u})$ 在 T 下是不变的. ■

下一个证明需要一个新记号. 设 U 和 W 都是 V 的子空间, 并且

$$V=U\oplus W.$$

每个向量 $\boldsymbol{v}\in V$ 都可以唯一地写成如下形式

$$\boldsymbol{v}=\boldsymbol{u}+\boldsymbol{w},$$

其中 $\boldsymbol{u}\in U$, $\boldsymbol{w}\in W$. 通过这个表示, 定义 $P_{U,W}\in\mathcal{L}(V)$ 如下

$$P_{U,W}\boldsymbol{v}=\boldsymbol{u}.$$

$P_{U,W}$ 通常称为到 U 上的带零空间 W 的**投影** (**projection**).

你应该验证, $P_{U,W}\boldsymbol{v}=\boldsymbol{v}$ 当且仅当 $\boldsymbol{v}\in U$. 在上面的表示中, 把 U 和 W 互换, 则有 $P_{W,U}\boldsymbol{v}=\boldsymbol{w}$. 因此, 对每个 $\boldsymbol{v}\in V$ 都有 $\boldsymbol{v}=P_{U,W}\boldsymbol{v}+P_{W,U}\boldsymbol{v}$. 你应该验证, $P_{U,W}^2=P_{U,W}$; 进一步, $\operatorname{range}P_{U,W}=U$, $\operatorname{null}P_{U,W}=W$.

我们已经见过 $\mathbf{R}^2$ 上的没有本征值的算子. 下面的定理证明了 $\mathbf{R}^3$ 上没有这样的例子.

5.26 定理: *在奇数维实向量空间上, 每个算子都有本征值.*

证明: 设 V 是奇数维实向量空间. 我们将对 V 的维数用归纳法 (步长为 2) 来证明 V 上算子都有本征值. 首先注意, 若 $\dim V=1$, 则结论显然成立.

现在假设 $\dim V$ 是大于 1 的奇数, 并假设结论对 $\dim V-2$ 维实向 量空间上的所有算子都成立. 设 $T\in\mathcal{L}(V)$, 则需要证明

T 有本征值. 倘若如此, 则证明结束. 否则, 由 5.24 知 V 有在 T 下不变的 2 维子空间 U. 由 2.13 知 V 有子空间 W 使得

$$V = U \oplus W.$$

因为 W 的维数为 $\dim V - 2$, 所以我们想对 $T|_W$ 用归纳法假设. 然而, W 可能不是在 T 下不变的, 也就是说 $T|_W$ 可能不是 W 上的算子. 为了得到 W 上的算子, 与投影 $P_{W,U}$ 复合. 具体地, 定义 $S \in \mathcal{L}(W)$ 如下

$$S\boldsymbol{w} = P_{W,U}(T\boldsymbol{w}), \quad \boldsymbol{w} \in W.$$

由归纳法假设, S 有一个本征值 λ. 往证这个 λ 是 T 的本征值.

设 $\boldsymbol{w} \in W$ 是 S 的相应于本征值 λ 的非零本征向量, 则 $(S - \lambda I)\boldsymbol{w} = \mathbf{0}$. 如果 $\boldsymbol{w}$ 是 T 的相应于本征值 λ 的本征向量, 则证明结束; 可惜情况未必如此. 因此, 我们将在 $U + \operatorname{span}(\boldsymbol{w})$ 中寻找 T 的本征向量. 为此, 考虑 $U + \operatorname{span}(\boldsymbol{w})$ 中的一个典型向量 $\boldsymbol{u} + a\boldsymbol{w}$, 其中 $\boldsymbol{u} \in U$, $a \in \mathbf{R}$, 则有

$$\begin{aligned}
(T - \lambda I)(\boldsymbol{u} + a\boldsymbol{w}) &= T\boldsymbol{u} - \lambda\boldsymbol{u} + a(T\boldsymbol{w} - \lambda\boldsymbol{w}) \\
&= T\boldsymbol{u} - \lambda\boldsymbol{u} + a(P_{U,W}(T\boldsymbol{w}) + P_{W,U}(T\boldsymbol{w}) - \lambda\boldsymbol{w}) \\
&= T\boldsymbol{u} - \lambda\boldsymbol{u} + a(P_{U,W}(T\boldsymbol{w}) + S\boldsymbol{w} - \lambda\boldsymbol{w}) \\
&= T\boldsymbol{u} - \lambda\boldsymbol{u} + aP_{U,W}(T\boldsymbol{w}).
\end{aligned}$$

注意到, 在上面等式的右端, $T\boldsymbol{u} \in U$ (因为 U 在 T 下是不变的), $\lambda\boldsymbol{u} \in U$ (因为 $\boldsymbol{u} \in U$), 并且 $aP_{U,W}(T\boldsymbol{w}) \in U$ (由 $P_{U,W}$ 的定义). 于是, $T - \lambda I$ 把 $U + \operatorname{span}(\boldsymbol{w})$ 映到 U 内. 因为 $U + \operatorname{span}(\boldsymbol{w})$ 的维数比 U 的维数大, 所以 $(T - \lambda I)|_{U+\operatorname{span}(\boldsymbol{w})}$ 不是单的 (参见 3.5). 也就是说, 有一个非零向量 $\boldsymbol{v} \in U + \operatorname{span}(\boldsymbol{w}) \subset V$ 使得 $(T - \lambda I)\boldsymbol{v} = \mathbf{0}$. 于是, T 有本征值. ■

习　题

1. 设 $T \in \mathcal{L}(V)$. 证明：若 V 的子空间 $U_1, \cdots, U_m$ 在 T 下都是不变的, 则 $U_1 + \cdots + U_m$ 在 T 下也是不变的.

2. 设 $T \in \mathcal{L}(V)$. 证明 V 的任意一族在 T 下不变的子空间的交也是在 T 下不变的.

3. 证明或举反例：若 V 的子空间 U 在 V 的每个算子下都是不变的, 则 $U = \{\mathbf{0}\}$ 或 $U = V$.

4. 设 $S, T \in \mathcal{L}(V)$ 使得 $ST = TS$. 证明：对每个 $\lambda \in F$, $\operatorname{null}(T - \lambda I)$ 在 S 下都是不变的.

5. 定义 $T \in \mathcal{L}(\mathbf{F}^2)$ 如下

$$T(w, z) = (z, w).$$

求 T 的所有本征值和本征向量.

6. 定义 $T \in \mathcal{L}(\mathbf{F}^3)$ 如下

$$T(z_1, z_2, z_3) = (2z_2, 0, 5z_3).$$

求 T 的所有本征值和本征向量.

7. 设 n 是正整数, 并设 $T \in \mathcal{L}(\mathbf{F}^n)$ 定义如下

$$T(x_1, \cdots, x_n) = (x_1 + \cdots + x_n, \cdots, x_1 + \cdots + x_n);$$

也就是说, 算子 T (关于标准基) 的矩阵的元素全是 1. 求 T 的所有本征值和本征向量.

8. 求如下定义的后向移位算子 $T \in \mathcal{L}(\mathbf{F}^\infty)$ 的所有本征值和本征向量,

$$T(z_1, z_2, z_3, \cdots) = (z_2, z_3, \cdots).$$

9. 设 $T \in \mathcal{L}(V)$, 并设 $\dim \operatorname{range} T = k$. 证明 T 最多有 $k + 1$ 个互不相同的本征值.

10. 设 $T \in \mathcal{L}(V)$ 可逆, 并设 $\lambda \in \mathbf{F} \backslash \{0\}$. 证明 λ 是 T 的本征值当且仅当 $\frac{1}{\lambda}$ 是 T^{-1} 的本征值.

11. 设 $S, T \in \mathcal{L}(V)$. 证明 ST 和 TS 有相同的本征值.

12. 设 $T \in \mathcal{L}(V)$, 并且 V 中每个向量都是 T 的本征向量. 证明 T 是恒等算子的标量倍.

13. 设 $T \in \mathcal{L}(V)$, 并且 V 的每个 $\dim V - 1$ 维子空间在 T 下都是不变的. 证明 T 是恒等算子的标量倍.

14. 设 $S, T \in \mathcal{L}(V)$, 并且 S 是可逆的. 证明: 若 $p \in \mathcal{P}(\mathbf{F})$ 是多项式, 则

$$p(STS^{-1}) = Sp(T)S^{-1}.$$

15. 设 $\mathbf{F} = \mathbf{C}$, $T \in \mathcal{L}(V)$, $p \in \mathcal{P}(\mathbf{C})$, $a \in \mathbf{C}$. 证明 a 是 $p(T)$ 的本征值当且仅当对于 T 的某个本征值 λ 有 $a = p(\lambda)$.

16. 证明前一个习题的结果对于 $\mathbf{R}$ 不成立.

17. 设 V 是复向量空间, 并设 $T \in \mathcal{L}(V)$. 证明: 对每个 $j = 1, \cdots, \dim V$, T 都有一个 j 维的不变子空间.

18. 给出一个可逆算子, 使得该算子关于某个基的矩阵的对角线上只有 0.

19. 给出一个不可逆算子, 使得该算子关于某个基的矩阵的对角线上的数都非零.

这两个习题表明, 去掉上三角矩阵的假设, 5.16 就不成立了.

20. 设 $T \in \mathcal{L}(V)$ 有 $\dim V$ 个互不相同的本征值, 并设 $S \in \mathcal{L}(V)$ 和 T 有相同的本征向量 (相应的本征值不必相同). 证明 $ST = TS$.

21. 设 $P \in \mathcal{L}(V)$, 并且 $P^2 = P$. 证明 $V = \operatorname{null} P \oplus \operatorname{range} P$.

22. 设 $V = U \oplus W$, 其中 U 和 W 都是 V 的非零子空间. 求 $P_{U,W}$ 的所有本征值和本征向量.

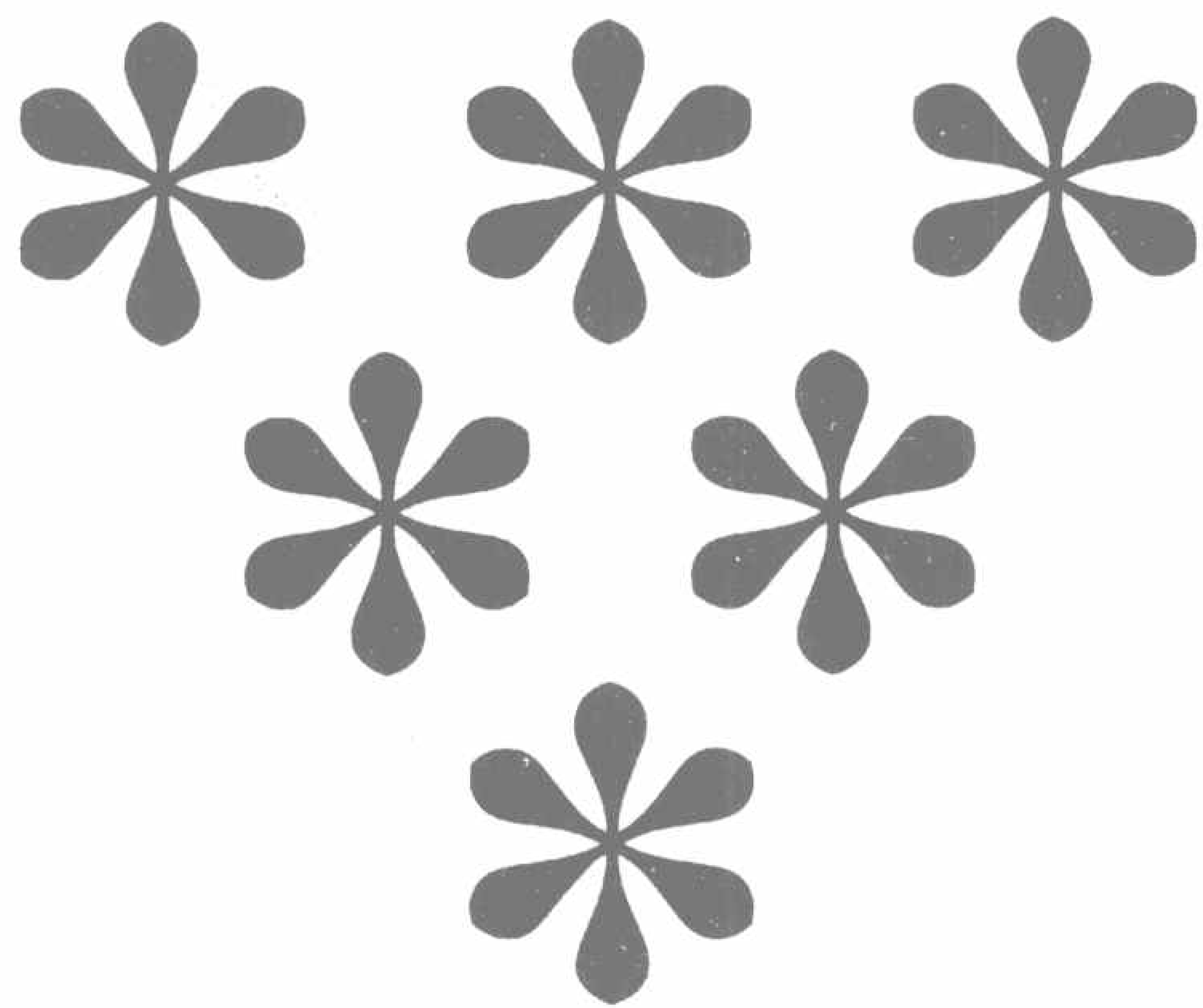
23. 给出一个没有 (实) 本征值的算子 $T \in \mathcal{L}(\mathbf{R}^4)$.

24. 设 V 是实向量空间, 并设 $T \in \mathcal{L}(V)$ 没有本征值. 证明 V 的每个在 T 下不变的子空间的维数都是偶数.

第6章 内积空间

我们在定义向量空间时推广了 $\mathbf{R}^2$ 和 $\mathbf{R}^3$ 的线性结构 (加法和标量乘法), 而忽略了其他的重要特征, 例如长度和角度的概念, 这些思想蕴含于我们现在要研究的内积的概念中.

$\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$.
V 是 $\mathbf{F}$ 上的有限维非零向量空间.



§6.1 内 积

为了说明引入内积概念的动机, 我们把 $\mathbf{R}^2$ 和 $\mathbf{R}^3$ 中的向量看作始于原点的箭头. $\mathbf{R}^2$ 或 $\mathbf{R}^3$ 中向量 $\boldsymbol{x}$ 的长度称为 $\boldsymbol{x}$ 的**范数 (norm)**, 记为 $\|\boldsymbol{x}\|$. 因此对 $\boldsymbol{x}=(x_1,x_2)\in\mathbf{R}^2$, 有 $\|\boldsymbol{x}\|=\sqrt{x_1^2+x_2^2}$, 见图 6-1.

如果我们把向量看作点, 而不是箭头, 那么 $\|\boldsymbol{x}\|$ 应该解释成从点 $\boldsymbol{x}$ 到原点的距离.

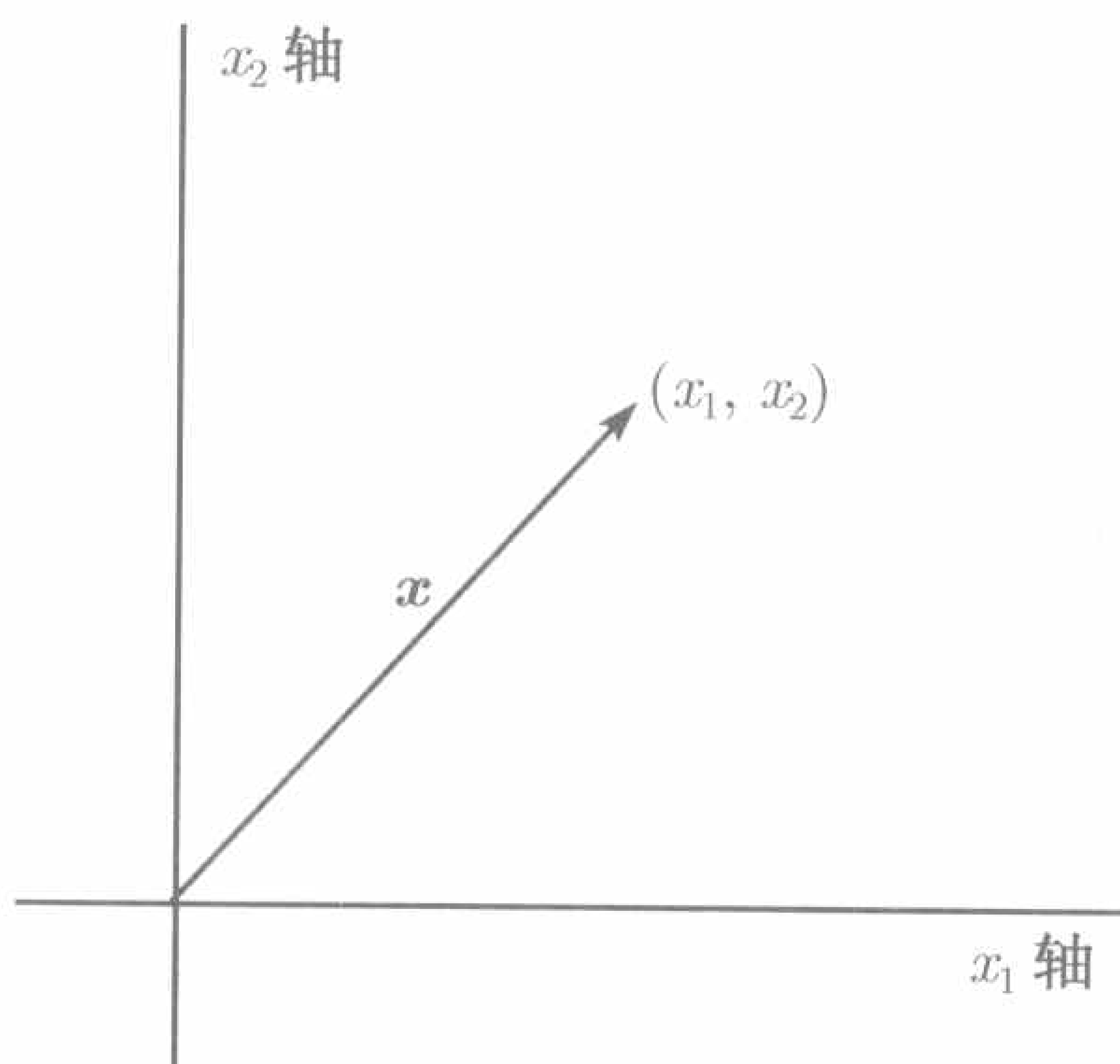


图 6-1 向量 $\boldsymbol{x}$ 的长度为 $\sqrt{x_1^2+x_2^2}$

类似地, 对 $\boldsymbol{x}=(x_1,x_2,x_3)\in\mathbf{R}^3$, 有 $\|\boldsymbol{x}\|=\sqrt{x_1^2+x_2^2+x_3^2}$. 虽然我们不能画出高维的图形, 但是范数在 $\mathbf{R}^n$ 上的推广是显然的: 定义 $\boldsymbol{x}=(x_1,\cdots,x_n)\in\mathbf{R}^n$ 的范数为

$$\|\boldsymbol{x}\|=\sqrt{x_1^2+\cdots+x_n^2}.$$

范数在 $\mathbf{R}^n$ 上不是线性的. 为了把线性引入讨论, 我们来介绍点积. 对于 $\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$, $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的**点积 (dot product)** 记为 $\boldsymbol{x}\cdot\boldsymbol{y}$, 定义为

$$\boldsymbol{x}\cdot\boldsymbol{y}=x_1y_1+\cdots+x_ny_n,$$

其中 $\boldsymbol{x}=(x_1,\cdots,x_n)$, $\boldsymbol{y}=(y_1,\cdots,y_n)$. 注意, $\mathbf{R}^n$ 中两个向量的点积是一个数, 而不是一个向量. 显然对所有 $\boldsymbol{x}\in\mathbf{R}^n$ 都有 $\boldsymbol{x}\cdot\boldsymbol{x}=\|\boldsymbol{x}\|^2$. 特别地, 对所有 $\boldsymbol{x}\in\mathbf{R}^n$ 都有 $\boldsymbol{x}\cdot\boldsymbol{x}\geqslant 0$, 其中等号成立当且仅当 $\boldsymbol{x}=\mathbf{0}$. 如果取定 $\boldsymbol{y}\in\mathbf{R}^n$, 则 $\boldsymbol{x}\in\mathbf{R}^n$ 到 $\boldsymbol{x}\cdot\boldsymbol{y}$ 的对应是从 $\mathbf{R}^n$ 到 $\mathbf{R}$ 的映射, 这个映射显然是线性的. 此外, 对所有 $\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$ 都有 $\boldsymbol{x}\cdot\boldsymbol{y}=\boldsymbol{y}\cdot\boldsymbol{x}$.

内积是点积的推广. 现在你可能会猜到, 内积是通过抽象上一段所讨论的点积的性质来定义的. 这个猜测对实向量空间是正确的. 但为了使定义对实向量空间和复向量空间都可用, 在给出定义之前, 我们需要考察复数的情形.

回想一下, 若 $\lambda = a + bi$, 其中 $a, b \in \mathbf{R}$, 那么 λ 的绝对值定义为

$$|\lambda| = \sqrt{a^2 + b^2},$$

λ 的复共轭定义为

$$\bar{\lambda} = a - bi,$$

并且等式

$$|\lambda|^2 = \lambda\bar{\lambda}$$

建立了这两个概念之间的联系 (绝对值和复共轭的定义和基本性质参见 69 页). 对 $\boldsymbol{z} = (z_1, \cdots, z_n) \in \mathbf{C}^n$, 定义 $\boldsymbol{z}$ 的范数为

$$\|\boldsymbol{z}\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

因为我们希望 $\|\boldsymbol{z}\|$ 是非负数, 所以上式中的绝对值是必要的. 注意到

$$\|\boldsymbol{z}\|^2 = z_1\overline{z_1} + \cdots + z_n\overline{z_n}.$$

我们想把 $\|\boldsymbol{z}\|^2$ 看作 $\boldsymbol{z}$ 与自身的内积, 就像在 $\mathbf{R}^n$ 中一样. 因此, 上面的等式提示我们, $\boldsymbol{w} = (w_1, \cdots, w_n) \in \mathbf{C}^n$ 与 $\boldsymbol{z}$ 的内积应该等于

$$w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

要是互换 $\boldsymbol{w}$ 和 $\boldsymbol{z}$ 的角色, 上面的表达式就要用它的复共轭来代替. 也就是说, $\boldsymbol{w}$ 和 $\boldsymbol{z}$ 的内积应该等于 $\boldsymbol{z}$ 和 $\boldsymbol{w}$ 的内积的复共轭. 有了这样的动机, 我们现在就可以定义 V 上的内积了, 这里 V 可以是实向量空间或者复向量空间.

V 上的**内积 (inner product)** 就是一个函数, 它把 V 中元素的每个有序对 $(\boldsymbol{u}, \boldsymbol{v})$ 都映成一个数 $\langle \boldsymbol{u}, \boldsymbol{v} \rangle \in \mathbf{F}$, 并且具有下列性质:

如果 z 是复数, 那么 $z \geqslant 0$ 的意思是 z 是实数并且是非负的.

正性 (positivity)

对所有 $v \in V$ 都有 $\langle v, v \rangle \geqslant 0$;

定性 (definiteness)

$\langle v, v \rangle = 0$ 当且仅当 $v = 0$;

第一个位置的加性 (additivity in first slot)

对所有 $u, v, w \in V$ 都有 $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$;

第一个位置的齐性 (homogeneity in first slot)

对所有 $a \in \mathbf{F}$, $v, w \in V$ 都有 $\langle av, w \rangle = a\langle v, w \rangle$;

共轭对称性 (conjugate symmetry)

对所有 $v, w \in V$ 都有 $\langle v, w \rangle = \overline{\langle w, v \rangle}$.

回想一下, 每个实数都等于它的复共轭. 因此在处理实向量空间时, 可以忽略上面最后一个条件中的复共轭, 而直接说: 对所有 $v, w \in V$ 都有 $\langle v, w \rangle = \langle w, v \rangle$.

内积空间 (inner-product space) 是带有内积的向量空间 V.

内积空间最重要的例子是 $\mathbf{F}^n$. 你应该验证, 可以如下定义 $\mathbf{F}^n$ 上的内积

如果我们处理的是 $\mathbf{R}^n$, 而不是 $\mathbf{C}^n$, 那么也可以忽略复共轭.

6.1 $$\langle (w_1, \cdots, w_n), (z_1, \cdots, z_n) \rangle = w_1\overline{z_1} + \cdots + w_n\overline{z_n},$$

称之为 $\mathbf{F}^n$ 上的**欧几里得内积** (Euclidean inner product), 正是它为内积的定义提供了动机. 在说 $\mathbf{F}^n$ 是内积空间时, 除非特别指明, 总假设采用的是欧几里得内积.

$\mathbf{F}^n$ 上除了欧几里得内积之外还有其他的内积. 例如, 若 $c_1, \cdots, c_n$ 是正数, 那么你应该验证, 可以如下定义 $\mathbf{F}^n$ 上的内积

$$\langle (w_1, \cdots, w_n), (z_1, \cdots, z_n) \rangle = c_1 w_1\overline{z_1} + \cdots + c_n w_n\overline{z_n}.$$

当然, 如果所有的 c 都等于 1, 则得到欧几里得内积.

再看内积空间的另一个例子, 考虑由系数在 $\mathbf{F}$ 中次数不超过 m 的所有多项式组成的向量空间 $\mathcal{P}_m(\mathbf{F})$. 你应该验证, 可以定义 $\mathcal{P}_m(\mathbf{F})$ 上的内积如下

6.2
$$\langle p,q\rangle=\int_0^1 p(x)\overline{q(x)}\mathrm{d}x.$$

若 $\mathbf{F}=\mathbf{R}$, 则不需要复共轭.

我们约定, 在本章的其余部分, V 是 $\mathbf{F}$ 上的有限维内积空间.

在内积的定义中, 第一个位置的加性和齐性这两个条件可以合并成为对第一个位置的线性要求. 更确切地说, 对每个固定的 $\boldsymbol{w}\in V$, 把 $\boldsymbol{v}$ 映成 $\langle\boldsymbol{v},\boldsymbol{w}\rangle$ 的函数是 V 到 $\mathbf{F}$ 的线性映射. 因为每个线性映射都把 $\mathbf{0}$ 映成 0, 所以一定有

$$\langle\mathbf{0},\boldsymbol{w}\rangle=0,\quad \boldsymbol{w}\in V.$$

因此 (由共轭对称性) 又有

$$\langle\boldsymbol{w},\mathbf{0}\rangle=0,\quad \boldsymbol{w}\in V.$$

在内积空间中, 像第一个位置一样, 第二个位置也具有加性. 证明:

$$\begin{aligned}\langle\boldsymbol{u},\boldsymbol{v}+\boldsymbol{w}\rangle&=\overline{\langle\boldsymbol{v}+\boldsymbol{w},\boldsymbol{u}\rangle}\\&=\overline{\langle\boldsymbol{v},\boldsymbol{u}\rangle+\langle\boldsymbol{w},\boldsymbol{u}\rangle}\\&=\overline{\langle\boldsymbol{v},\boldsymbol{u}\rangle}+\overline{\langle\boldsymbol{w},\boldsymbol{u}\rangle}\\&=\langle\boldsymbol{u},\boldsymbol{v}\rangle+\langle\boldsymbol{u},\boldsymbol{w}\rangle;\end{aligned}$$

这里 $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in V$.

在内积空间中, 第二个位置具有共轭齐性, 即对所有的标量 $a\in\mathbf{F}$ 都有 $\langle\boldsymbol{u},a\boldsymbol{v}\rangle=\bar{a}\langle\boldsymbol{u},\boldsymbol{v}\rangle$. 证明:

$$\begin{aligned}\langle\boldsymbol{u},a\boldsymbol{v}\rangle&=\overline{\langle a\boldsymbol{v},\boldsymbol{u}\rangle}\\&=\overline{a\langle\boldsymbol{v},\boldsymbol{u}\rangle}\\&=\bar{a}\overline{\langle\boldsymbol{v},\boldsymbol{u}\rangle}\\&=\bar{a}\langle\boldsymbol{u},\boldsymbol{v}\rangle;\end{aligned}$$

这里 $a\in\mathbf{F}$, $\boldsymbol{u},\boldsymbol{v}\in V$. 注意, 在实向量空间中, 共轭齐性与齐性是相同的.

§6.2 范 数

对于 $v \in V$, v 的范数, 记为 $\|v\|$, 定义为

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

例如, 若 $(z_1, \cdots, z_n) \in \mathbf{F}^n$ (取欧几里得内积), 那么

$$\|(z_1, \cdots, z_n)\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

再看另一个例子, 如果 $p \in \mathcal{P}_m(\mathbf{F})$ (取 6.2 所给出的内积), 那么

$$\|p\| = \sqrt{\int_0^1 |p(x)|^2 \mathrm{d}x}.$$

注意到 $\|v\| = 0$ 当且仅当 $v = \mathbf{0}$ (因为 $\langle v, v\rangle = 0$ 当且仅当 $v = \mathbf{0}$). 范数的另一个简单性质是, 对所有 $a \in \mathbf{F}, v \in V$ 都有 $\|av\| = |a|\,\|v\|$. 证明如下:

$$\begin{aligned}\|av\|^2 &= \langle av, av\rangle \\ &= a\langle v, av\rangle \\ &= a\bar{a}\langle v, v\rangle \\ &= |a|^2\|v\|^2;\end{aligned}$$

现在取平方根即得要证的等式. 这个证明阐明了一个普遍原理: 处理范数的平方通常比直接处理范数容易.

对于两个向量 $u, v \in V$, 如果 $\langle u, v\rangle = 0$, 则称 u 和 v 是**正交的** (**orthogonal**). 注意, 向量的次序是无关紧要的, 因为 $\langle u, v\rangle = 0$ 当且仅当 $\langle v, u\rangle = 0$. 有时候我们说 u 正交于 v, 而不说 u 和 v 是正交的. 显然 $\mathbf{0}$ 正交于每个向量. 进一步, $\mathbf{0}$ 是唯一正交于自身的向量.

有些数学家采用**垂直** (**perpendicular**) 这样的术语, 意思与正交相同.

正交一词来源于希腊语 orthogonios, 意思是直角的.

下一个定理是对于 $V = \mathbf{R}^2$ 的特殊情况, 它已经有 3000 多年历史了.

6.3 勾股定理 (也称毕达哥拉斯定理 (Pythagorean Theorem)): 如果 u, v 是 V 中的正交向量, 那么

6.4 $$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

证明： 设 $\boldsymbol{u}, \boldsymbol{v}$ 是 V 中的正交向量, 则

$$\begin{aligned}\|\boldsymbol{u}+\boldsymbol{v}\|^2 &= \langle \boldsymbol{u}+\boldsymbol{v}, \boldsymbol{u}+\boldsymbol{v}\rangle \\ &= \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + \langle \boldsymbol{u}, \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{u}\rangle \\ &= \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2,\end{aligned}$$

证毕. ■

勾股定理的证明表明, 6.4 成立当且仅当 $\langle \boldsymbol{u}, \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{u}\rangle = 0$, 即 $2\mathrm{Re}\,\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$. 因此勾股定理的逆命题在实内积空间中成立.

设 $\boldsymbol{u}, \boldsymbol{v} \in V$. 我们想把 $\boldsymbol{u}$ 写成 $\boldsymbol{v}$ 的标量倍加上一个正交于 $\boldsymbol{v}$ 的向量, 如图 6-2 所示.

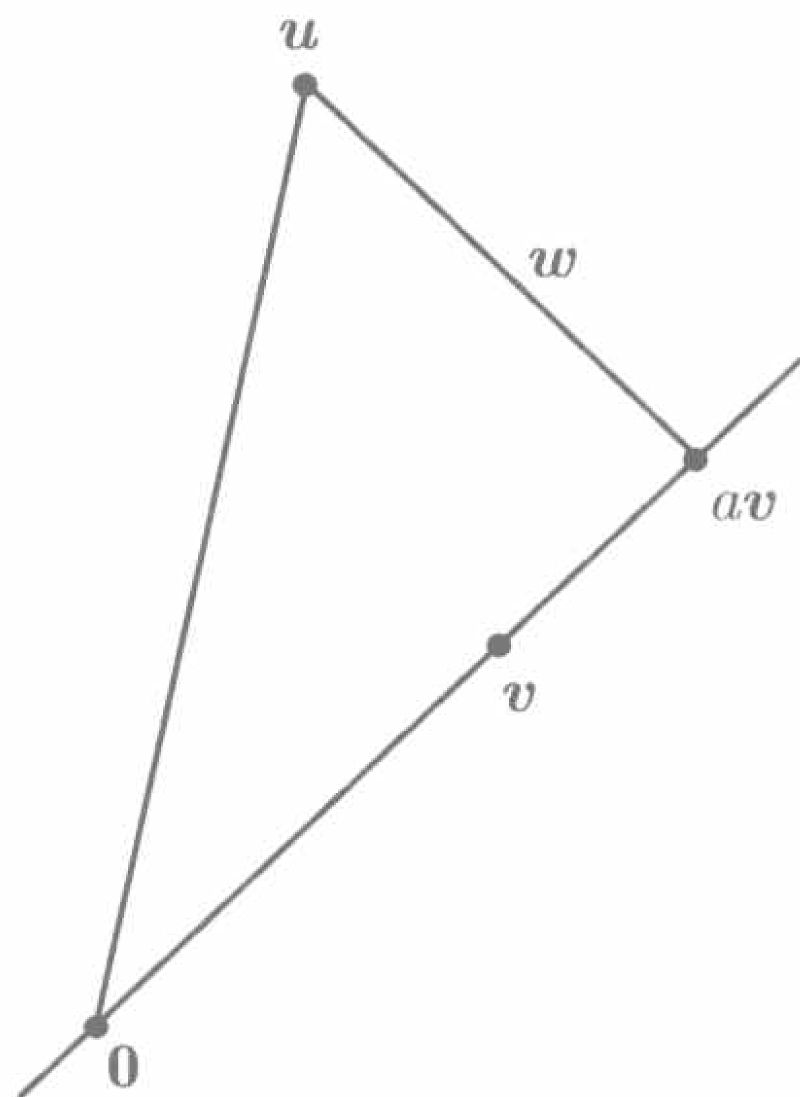


图 6-2 正交分解

为了揭示如何将 $\boldsymbol{u}$ 写成 $\boldsymbol{v}$ 的标量倍加上一个正交于 $\boldsymbol{v}$ 的向量, 令 $a \in \mathbf{F}$ 表示一个标量, 则

$$\boldsymbol{u} = a\boldsymbol{v} + (\boldsymbol{u} - a\boldsymbol{v}).$$

因此需要选取 a 使得 $\boldsymbol{v}$ 正交于 $(\boldsymbol{u} - a\boldsymbol{v})$. 也就是说, 我们想要

$$0 = \langle \boldsymbol{u} - a\boldsymbol{v}, \boldsymbol{v}\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle - a\|\boldsymbol{v}\|^2.$$

上面的等式表明 a 应取成 $\langle \boldsymbol{u}, \boldsymbol{v}\rangle / \|\boldsymbol{v}\|^2$ (为了避免分母为 0, 假设 $\boldsymbol{v} \neq \boldsymbol{0}$), 从而

6.5
$$\boldsymbol{u} = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{v}\|^2}\boldsymbol{v} + \left(\boldsymbol{u} - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{v}\|^2}\boldsymbol{v}\right).$$

你应该验证, 若 $\boldsymbol{v} \neq \boldsymbol{0}$, 则上式就把 $\boldsymbol{u}$ 写成了 $\boldsymbol{v}$ 的标量倍加上一个正交于 $\boldsymbol{v}$ 的向量.

下一个定理的证明要用到等式 6.5, 这个定理是数学中最重要的不等式之一.

1821 年, 法国数学家柯西 (Augustin - Louis Cauchy) 证明了这个不等式对 6.1 定义的内积成立. 1886 年, 德国数学家施瓦茨 (Herman Schwarz) 证明了这个不等式对 6.2 定义的内积成立.

6.6 柯西–施瓦茨不等式 (Cauchy-Schwarz Inequality): 若 $\boldsymbol{u}, \boldsymbol{v} \in V$, 则

6.7 $$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leqslant \|\boldsymbol{u}\|\|\boldsymbol{v}\|,$$

而且其中的等号成立当且仅当 $\boldsymbol{u}, \boldsymbol{v}$ 之一是另一个的标量倍.

证明: 设 $\boldsymbol{u}, \boldsymbol{v} \in V$. 若 $\boldsymbol{v} = \boldsymbol{0}$, 则 6.7 的两端都等于 0, 不等式成立. 因此可以假设 $\boldsymbol{v} \neq \boldsymbol{0}$. 考虑正交分解

$$\boldsymbol{u} = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{v}\|^2}\boldsymbol{v} + \boldsymbol{w},$$

其中 $\boldsymbol{w}$ 正交于 $\boldsymbol{v}$ (这里 $\boldsymbol{w}$ 等于 6.5 右端的第二项). 由勾股定理,

$$\begin{aligned}\|\boldsymbol{u}\|^2 &= \left\|\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{v}\|^2}\boldsymbol{v}\right\|^2 + \|\boldsymbol{w}\|^2 \\ &= \frac{|\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2}{\|\boldsymbol{v}\|^2} + \|\boldsymbol{w}\|^2 \\ &\geqslant \frac{|\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2}{\|\boldsymbol{v}\|^2}. \qquad \textbf{6.8}\end{aligned}$$

不等式两端都乘以 $\|\boldsymbol{v}\|^2$, 再取平方根即得柯西–施瓦茨不等式 6.7.

从柯西–施瓦茨不等式的证明可以看出 6.7 是等式当且仅当 6.8 是等式. 显然, 这一条件成立当且仅当 $\boldsymbol{w} = \boldsymbol{0}$. 而 $\boldsymbol{w} = \boldsymbol{0}$ 当且仅当 $\boldsymbol{u}$ 是 $\boldsymbol{v}$ 的标量倍 (参见 6.5). 因此, 柯西–施瓦茨不等式是等式当且仅当 $\boldsymbol{u}$ 是 $\boldsymbol{v}$ 的标量倍或 $\boldsymbol{v}$ 是 $\boldsymbol{u}$ 的标量倍 (或者二者都成立; 这样措辞是为了涵盖 $\boldsymbol{u}$ 或 $\boldsymbol{v}$ 等于$\boldsymbol{0}$的情况). ■

下一个结果称为三角不等式, 因为它的几何解释为: 三角形中任意一边的长度小于另外两边的长度之和.

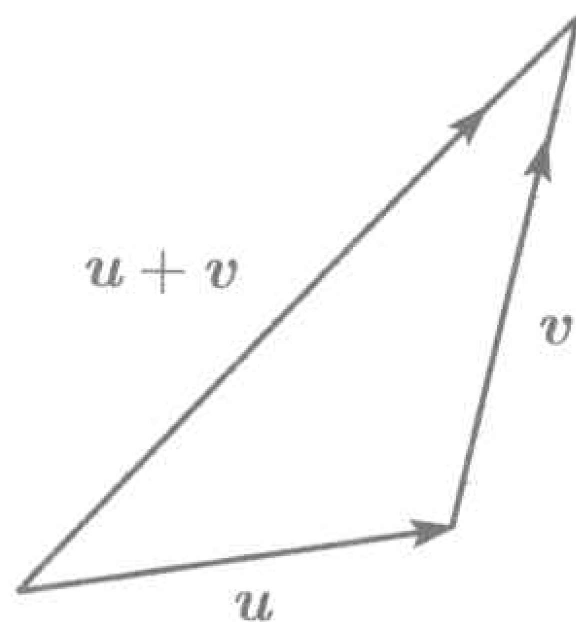


图 6-3 三角不等式

6.9 三角不等式 (Triangle Inequality)： 若 $\boldsymbol{u}, \boldsymbol{v} \in V$, 则

三角不等式可以用来证明，两点之间的最短路线是直线.

6.10 $$\|\boldsymbol{u}+\boldsymbol{v}\| \leqslant \|\boldsymbol{u}\|+\|\boldsymbol{v}\|,$$

而且其中的等号成立当且仅当 $\boldsymbol{u}, \boldsymbol{v}$ 之一是另一个的非负标量倍.

证明： 设 $\boldsymbol{u}, \boldsymbol{v} \in V$, 则

$$\begin{aligned}
\|\boldsymbol{u}+\boldsymbol{v}\|^2 &= \langle \boldsymbol{u}+\boldsymbol{v}, \boldsymbol{u}+\boldsymbol{v}\rangle \\
&= \langle \boldsymbol{u}, \boldsymbol{u}\rangle + \langle \boldsymbol{v}, \boldsymbol{v}\rangle + \langle \boldsymbol{u}, \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{u}\rangle \\
&= \langle \boldsymbol{u}, \boldsymbol{u}\rangle + \langle \boldsymbol{v}, \boldsymbol{v}\rangle + \langle \boldsymbol{u}, \boldsymbol{v}\rangle + \overline{\langle \boldsymbol{u}, \boldsymbol{v}\rangle} \\
&= \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2\mathrm{Re}\langle \boldsymbol{u}, \boldsymbol{v}\rangle \\
&\leqslant \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2|\langle \boldsymbol{u}, \boldsymbol{v}\rangle| \qquad \textbf{6.11} \\
&\leqslant \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2\|\boldsymbol{u}\|\|\boldsymbol{v}\| \\
&= (\|\boldsymbol{u}\| + \|\boldsymbol{v}\|)^2, \qquad \textbf{6.12}
\end{aligned}$$

其中 6.12 由柯西–施瓦茨不等式 (6.6) 得到. 上面不等式两端取平方根即得三角不等式 6.10.

上面的证明表明，三角不等式 6.10 是等式当且仅当 6.11 和 6.12 都是等式. 因此，三角不等式 6.10 是等式当且仅当

6.13 $$\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \|\boldsymbol{u}\|\|\boldsymbol{v}\|.$$

你应该验证，如果 $\boldsymbol{u}, \boldsymbol{v}$ 之一是另一个的非负标量倍，那么 6.13 成立. 反之，设 6.13 成立，则由柯西–施瓦茨不等式 (6.6) 等号成立的条件可知，$\boldsymbol{u}, \boldsymbol{v}$ 之一必定是另一个的标量倍，再由 6.13 知这个标量显然是非负的. ■

下一个结果称为平行四边形等式，因为它的几何解释为：在任意平行四边形中，对角线长度的平方和等于四条边长度的平方和.

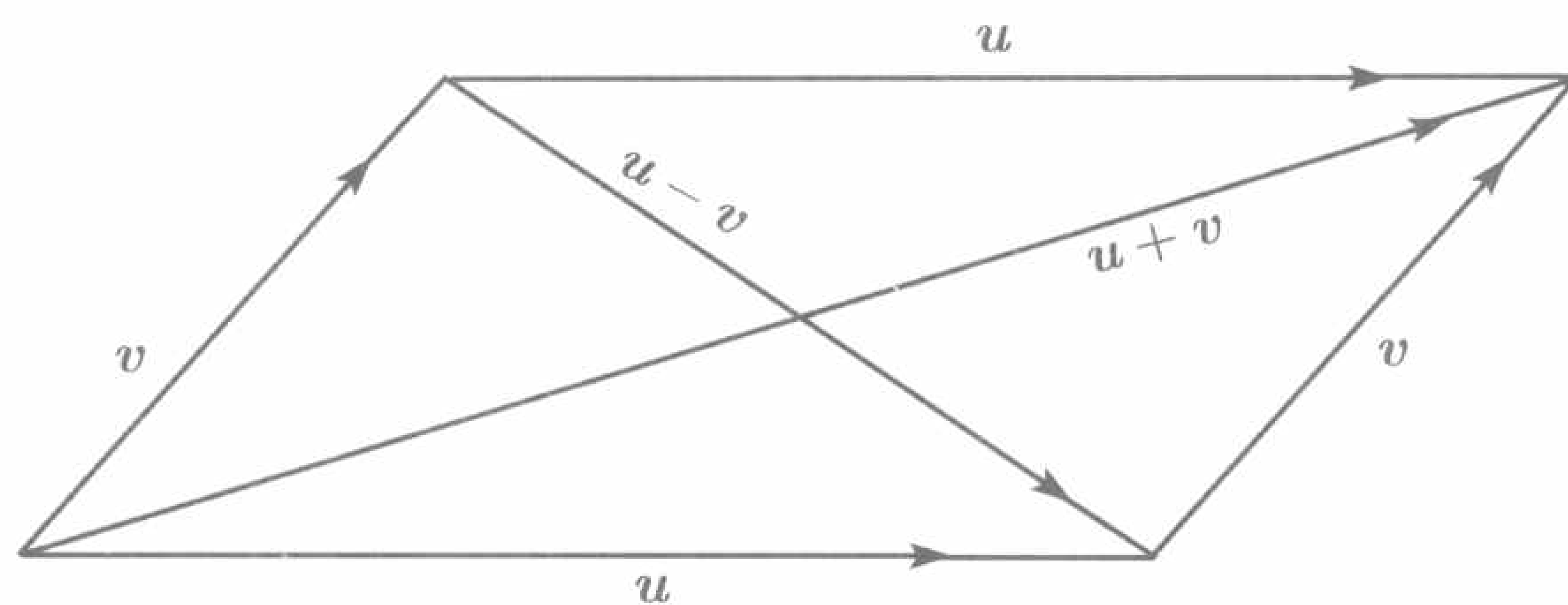


图 6-4 平行四边形等式

6.14 平行四边形等式 (Parallelogram Equality)：若 $\boldsymbol{u}, \boldsymbol{v} \in V$，则

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2+\|\boldsymbol{u}-\boldsymbol{v}\|^2=2(\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2).$$

证明：设 $\boldsymbol{u}, \boldsymbol{v} \in V$, 则

$$\begin{aligned}
\|\boldsymbol{u}+\boldsymbol{v}\|^2+\|\boldsymbol{u}-\boldsymbol{v}\|^2 &= \langle \boldsymbol{u}+\boldsymbol{v}, \boldsymbol{u}+\boldsymbol{v}\rangle+\langle \boldsymbol{u}-\boldsymbol{v}, \boldsymbol{u}-\boldsymbol{v}\rangle \\
&= \|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2+\langle \boldsymbol{u}, \boldsymbol{v}\rangle+\langle \boldsymbol{v}, \boldsymbol{u}\rangle \\
&\quad +\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2-\langle \boldsymbol{u}, \boldsymbol{v}\rangle-\langle \boldsymbol{v}, \boldsymbol{u}\rangle \\
&= 2(\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2),
\end{aligned}$$

证毕. ■

§6.3 规范正交基

如果一个向量组中的向量两两正交, 并且每个向量的范数都为 1, 则称这个向量组是**规范正交的 (orthonormal)**. 也就是说, V 中向量组 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 是规范正交的当且仅当 $\langle \boldsymbol{e}_j, \boldsymbol{e}_k\rangle$ 在 $j \neq k$ 时等于 0, 而在 $j=k$ 时等于 1 $(j, k=1, \cdots, m)$. 例如, $\mathbf{F}^n$ 中的标准基是规范正交的. 下一个命题表明规范正交组是很容易处理的.

6.15 命题：如果 V 中向量组 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 是规范正交的, 那么

$$\|a_1\boldsymbol{e}_1+\cdots+a_m\boldsymbol{e}_m\|^2=|a_1|^2+\cdots+|a_m|^2,$$

其中 $a_1, \cdots, a_m \in \mathbf{F}$.

证明：因为每个 $\boldsymbol{e}_j$ 的范数都为 1, 通过反复使用勾股定理 (6.3) 容易得证. ■

现在我们有如下简单而重要的推论.

6.16 推论: 每个规范正交向量组都是线性无关的.

证明: 假设 V 中向量组 $(e_1, \cdots, e_m)$ 是规范正交的, 并且 $a_1, \cdots, a_m \in \mathbf{F}$ 使得

$$a_1 e_1 + \cdots + a_m e_m = 0.$$

那么 $|a_1|^2 + \cdots + |a_m|^2 = 0$ (由 6.15), 即所有的 a_j 都是 0. ■

V 的**规范正交基** (**orthonormal basis**) 是 V 中的一个规范正交向量组构成的基. 例如, 标准基是 $\mathbf{F}^n$ 的规范正交基. V 中每个长度为 $\dim V$ 的规范正交向量组自然是 V 的规范正交基 (证明: 由推论 6.16, 任何这样的组都一定是线性无关的; 又因为它具有适当的长度, 所以一定是基 —— 参见 2.17). 为了阐明这一原理, 考虑 $\mathbf{R}^4$ 中的四个向量构成的向量组:

$$\left(\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right), \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right), \left(-\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)\right).$$

容易证明这个组是规范正交的 (证明它!); 又因为这个组含有四个向量, 所以它一定是规范正交基.

一般来说, 给定 V 的基 $(e_1, \cdots, e_n)$ 和向量 $v \in V$, 必有标量 $a_1, \cdots, a_n$ 使得

$$v = a_1 e_1 + \cdots + a_n e_n,$$

但要找出这些 a_j 可能很困难. 然而, 下一个定理表明这对于规范正交基却很容易.

6.17 定理: 设 $(e_1, \cdots, e_n)$ 是 V 的规范正交基, 则对每个 $v \in V$, 都有

规范正交基的重要性主要源于这个定理.

6.18 $$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

而且

6.19 $$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

证明：设 $v \in V$. 因为 $(e_1, \cdots, e_n)$ 是 V 的基, 所以存在标量 $a_1, \cdots, a_n$ 使得

$$v = a_1 e_1 + \cdots + a_n e_n.$$

等式两端都与 e_j 做内积可得 $\langle v, e_j \rangle = a_j$. 因此 6.18 成立. 由 6.18 和 6.15 显然可得 6.19. ■

既然我们理解了规范正交基的作用, 那么怎么找到它们呢? 例如, 对于由区间 $[0,1]$ 上的定积分所定义的内积 (参见 6.2), $\mathcal{P}_m(\mathbf{F})$ 有正交基吗? 我们即将看到, 下一个结果将会给出这些问题的答案. 下面证明中所用到的算法称为**格拉姆–施密特过程 (Gram-Schmidt procedure)**. 它所给出的方法, 可以把一个线性无关组转化成与原来的组有相同张成的规范正交组.

丹麦数学家格拉姆 (Jorgen Gram, 1850—1916) 和德国数学家施密特 (Erhard Schmidt, 1876—1959) 把这一算法推广开来, 用以构造规范正交基.

6.20 格拉姆–施密特过程 (Gram-Schmidt procedure)：如果 $(v_1, \cdots, v_m)$ 是 V 中的线性无关向量组, 则 V 有规范正交向量组 $(e_1, \cdots, e_m)$ 使得

6.21
$$\operatorname{span}(v_1, \cdots, v_j) = \operatorname{span}(e_1, \cdots, e_j), \quad j = 1, \cdots, m.$$

证明：假设 $(v_1, \cdots, v_m)$ 是 V 中的线性无关向量组. 要构造这些 e, 首先令 $e_1 = v_1 / \|v_1\|$. 它满足 $j = 1$ 时的 6.21. 我们将如下归纳地选取 $e_2, \cdots, e_m$. 假设 $j > 1$, 并且已经选取了规范正交组 $(e_1, \cdots, e_{j-1})$ 使得

6.22
$$\operatorname{span}(v_1, \cdots, v_{j-1}) = \operatorname{span}(e_1, \cdots, e_{j-1}).$$

令

6.23
$$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{\| v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1} \|}.$$

注意到 $v_j \notin \operatorname{span}(v_1, \cdots, v_{j-1})$ (因为 $(v_1, \cdots, v_m)$ 是线性无关的), 则 $v_j \notin \operatorname{span}(e_1, \cdots, e_{j-1})$, 因此上式中的分母不为 0, 于是 e_j 定义合理. 用一个向量除以它的范数就得到一个范数为 1 的新向量; 因此 $\|e_j\| = 1$.

设 $1 \leqslant k < j$, 则

$$\begin{aligned}\langle \boldsymbol{e}_j, \boldsymbol{e}_k\rangle &= \left\langle \frac{\boldsymbol{v}_j - \langle \boldsymbol{v}_j, \boldsymbol{e}_1\rangle \boldsymbol{e}_1 - \cdots - \langle \boldsymbol{v}_j, \boldsymbol{e}_{j-1}\rangle \boldsymbol{e}_{j-1}}{\|\boldsymbol{v}_j - \langle \boldsymbol{v}_j, \boldsymbol{e}_1\rangle \boldsymbol{e}_1 - \cdots - \langle \boldsymbol{v}_j, \boldsymbol{e}_{j-1}\rangle \boldsymbol{e}_{j-1}\|}, \boldsymbol{e}_k \right\rangle \\ &= \frac{\langle \boldsymbol{v}_j, \boldsymbol{e}_k\rangle - \langle \boldsymbol{v}_j, \boldsymbol{e}_k\rangle}{\|\boldsymbol{v}_j - \langle \boldsymbol{v}_j, \boldsymbol{e}_1\rangle \boldsymbol{e}_1 - \cdots - \langle \boldsymbol{v}_j, \boldsymbol{e}_{j-1}\rangle \boldsymbol{e}_{j-1}\|} \\ &= 0.\end{aligned}$$

因此 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_j)$ 是规范正交组.

由 6.23 得 $\boldsymbol{v}_j \in \operatorname{span}(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_j)$. 再结合 6.22 得

$$\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_j) \subset \operatorname{span}(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_j).$$

上面的这两个组都是线性无关的 (这些 $\boldsymbol{v}$ 的无关性是根据假设得到的, 这些 $\boldsymbol{e}$ 的无关性是由规范正交性和 6.16 得到的). 因此上面两个子空间的维数都为 j, 从而一定相等. ■

现在我们可以解决规范正交基的存在性问题.

6.24 推论: *每个有限维内积空间都有规范正交基.*

在这个推论之前我们对内积空间的讨论都不需要标准假设: V 是有限维的.

证明: 取 V 的一个基. 对它应用格拉姆–施密特过程 (6.20), 则得到一个规范正交组. 这个规范正交组是线性无关的 (由 6.16) 并且它的张成等于 V, 因此是 V 的规范正交基. ■

有时候我们必须知道, 不仅规范正交基是存在的, 而且任何规范正交组都可以扩充成规范正交基. 在下一个推论中, 格拉姆–施密特过程表明这种扩充总是可能的.

6.25 推论: *V 中每个规范正交向量组都可以扩充成 V 的规范正交基.*

证明: 设 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 是 V 中的规范正交向量组, 则 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 是线性无关的 (由 6.16), 从而可以扩充成 V 的基 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ (参见 2.12). 现在对 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$

应用格拉姆–施密特过程 (6.20), 得到规范正交组

6.26 $$(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m, \boldsymbol{f}_1, \cdots, \boldsymbol{f}_n);$$

这里格拉姆–施密特过程保持前 m 个向量不变, 因为它们已经是规范正交的. 显然 6.26 是 V 的规范正交基, 因为它是线性无关的 (由 6.16) 并且它的张成等于 V. 所以我们已经将 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 扩充成了 V 的规范正交基. ■

回想一下, 如果一个矩阵对角线下方的所有元素都等于 0, 则称这个矩阵是上三角的. 也就是说, 一个上三角矩阵形如:

$$\begin{bmatrix} * & & * \\ & \ddots & \\ 0 & & * \end{bmatrix}.$$

在上一章我们证明了, 如果 V 是复向量空间, 那么对 V 上的每个算子都存在一个基使得该算子关于此基具有上三角矩阵 (参见 5.13). 既然现在讨论的是内积空间, 我们想知道何时存在规范正交基使得算子关于此基有上三角矩阵. 下一个推论表明, 能使得 T 具有上三角矩阵的基的存在性蕴含着具有同样性质的规范正交基的存在性. 这一结果在实向量空间和复向量空间上都成立 (但在实向量空间上, 这个假设只对某些算子成立).

6.27 推论: *假设 $T \in \mathcal{L}(V)$. 如果 T 关于 V 的某个基具有上三角矩阵, 那么 T 关于 V 的某个规范正交基也具有上三角矩阵.*

证明: 设 T 关于 V 的基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 具有上三角矩阵, 那么对每个 $j = 1, \cdots, n$, 都有 $\operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_j)$ 在 T 下是不变的 (参见 5.12).

对 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 应用格拉姆–施密特过程, 得到了 V 的规范正交基 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$. 因为对每个 j 都有

$$\operatorname{span}(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_j) = \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_j)$$

(参见 6.21), 所以, 对每个 $j = 1, \cdots, n$, $\operatorname{span}(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_j)$ 在 T 下

都是不变的. 因此, 由 5.12, T 关于规范正交基 $(e_1, \cdots, e_n)$ 具有上三角矩阵. ■

下一个结果是上一个推论的重要应用.

6.28 推论: 设 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$, 则 T 关于 V 的某个规范正交基具有上三角矩阵.

这个结果有时称为舒尔定理. 德国数学家舒尔 (Issai Schur) 在 1909 年发表了这个结果的第一个证明.

证明: 由 5.13 和 6.27 立即可得. ■

§6.4 正交投影与极小化问题

如果 U 是 V 的子集, 那么 U 的**正交补 (orthogonal complement)**, 记为 $U^\perp$, 是由 V 中与 U 的每个向量都正交的那些向量组成的集合:

$$U^\perp = \{v \in V : \langle v, u \rangle = 0, u \in U\}.$$

你应该验证, $U^\perp$ 总是 V 的子空间, $V^\perp = \{0\}$, 并且 $\{0\}^\perp = V$. 注意到, 若 $U_1 \subset U_2$, 则 $U_1^\perp \supset U_2^\perp$.

回想一下, 如果 U_1, U_2 都是 V 的子空间, 并且 V 中每个元素都可以唯一地写成 U_1 中的一个向量与 U_2 中的一个向量的和, 那么 V 是 U_1 和 U_2 的直和 (记为 $V = U_1 \oplus U_2$). 下一个定理表明, 由内积空间的每个子空间都可以得到整个空间的一个自然的直和分解.

6.29 定理: 如果 U 是 V 的子空间, 那么

$$V = U \oplus U^\perp.$$

证明: 设 U 是 V 的子空间. 首先证明

6.30 $$V = U + U^\perp.$$

为此, 设 $v \in V$, 并且 $(e_1, \cdots, e_m)$ 为 U 的一个规范正交基. 显然

6.31
$$v=\underbrace{\langle v,e_1\rangle e_1+\cdots+\langle v,e_m\rangle e_m}_{u}+\underbrace{v-\langle v,e_1\rangle e_1-\cdots-\langle v,e_m\rangle e_m}_{w}.$$

显然 $u\in U$. 因为 $(e_1,\cdots,e_m)$ 是一个规范正交组, 所以对每个 j 都有

$$\begin{aligned}\langle w,e_j\rangle&=\langle v,e_j\rangle-\langle v,e_j\rangle\\&=0.\end{aligned}$$

因此, w 正交于 $\mathrm{span}(e_1,\cdots,e_m)$ 中的每个向量. 即 $w\in U^\perp$. 于是 $v=u+w$, 其中 $u\in U$, $w\in U^\perp$, 这就证明了 6.30.

若 $v\in U\cap U^\perp$, 则 v (包含于 U) 正交于 U 中每个向量 (其中包括 v 本身), 从而 $\langle v,v\rangle=0$, 即 $v=0$. 因此

6.32 $$U\cap U^\perp=\{0\}.$$

由 6.30 和 6.32 可得 $V=U\oplus U^\perp$ (参见 1.9). ■

下一个推论是定理 6.29 的重要结果.

6.33 推论: 如果 U 是 V 的子空间, 那么

$$U=(U^\perp)^\perp.$$

证明: 设 U 是 V 的子空间. 首先证明

6.34 $$U\subset(U^\perp)^\perp.$$

为此, 设 $u\in U$, 则对每个 $v\in U^\perp$ 都有 $\langle u,v\rangle=0$ (根据 $U^\perp$ 的定义). 因为 u 正交于 $U^\perp$ 中的每个向量, 所以 $u\in(U^\perp)^\perp$, 这就证明了 6.34.

要证明另一个方向的包含关系, 设 $v\in(U^\perp)^\perp$. 由 6.29 可得 $v=u+w$, 其中 $u\in U$, $w\in U^\perp$. 从而 $v-u=w\in U^\perp$. 因为 $v\in(U^\perp)^\perp$, 并且 $u\in(U^\perp)^\perp$ (由 6.34), 所以 $v-u\in(U^\perp)^\perp$. 因此 $v-u\in U^\perp\cap(U^\perp)^\perp$, 由此知 $v-u$ 正交于自身, 从而 $v-u=0$, 即 $v=u$, 于是 $v\in U$. 因此 $(U^\perp)^\perp\subset U$, 再结合 6.34 就完成了证明. ■

设 U 是 V 的子空间. 6.29 所给出的分解 $V = U \oplus U^\perp$ 表明, 每个向量 $\boldsymbol{v} \in V$ 都可以唯一地写成

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w},$$

其中 $\boldsymbol{u} \in U$, $\boldsymbol{w} \in U^\perp$. 利用此分解来定义 V 上的算子 P_U, 称为 V 到 U 上的**正交投影 (orthogonal projection)**: 对 $\boldsymbol{v} \in V$, 定义 $P_U\boldsymbol{v}$ 为上面分解中的向量 $\boldsymbol{u}$. 采用所引进的记号则有 $P_U = P_{U,U^\perp}$. 你应该验证, $P_U \in \mathcal{L}(V)$, 并且具有以下性质:

- range $P_U = U$;
- null $P_U = U^\perp$;
- 对每个 $\boldsymbol{v} \in V$ 都有 $\boldsymbol{v} - P_U\boldsymbol{v} \in U^\perp$;
- ${P_U}^2 = P_U$;
- 对每个 $\boldsymbol{v} \in V$ 都有 $\|P_U\boldsymbol{v}\| \leqslant \|\boldsymbol{v}\|$.

进一步, 由 6.29 的证明中所使用的分解 6.31 可知, 如果 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 是 U 的规范正交基, 那么对每个 $\boldsymbol{v} \in V$ 都有

6.35 $$P_U\boldsymbol{v} = \langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \cdots + \langle \boldsymbol{v}, \boldsymbol{e}_m \rangle \boldsymbol{e}_m.$$

经常会遇到这样的问题: 给定 V 的子空间 U 和点 $\boldsymbol{v} \in V$, 求点 $\boldsymbol{u} \in U$ 使得 $\|\boldsymbol{v} - \boldsymbol{u}\|$ 最小. 下一个命题表明, 通过取 $\boldsymbol{u} = P_U\boldsymbol{v}$ 就可以解决这个极小化问题.

6.36 命题: 设 U 是 V 的子空间, 并且 $\boldsymbol{v} \in V$, 则

$$\|\boldsymbol{v} - P_U\boldsymbol{v}\| \leqslant \|\boldsymbol{v} - \boldsymbol{u}\|, \quad \boldsymbol{u} \in U.$$

进一步, 若 $\boldsymbol{u} \in U$ 使得上面的不等式是等式, 则 $\boldsymbol{u} = P_U\boldsymbol{v}$.

极小化问题求解的显著简化导致了内积空间在纯数学之外的很多应用.

证明: 设 $\boldsymbol{u} \in U$, 则

6.37 $$\|\boldsymbol{v} - P_U\boldsymbol{v}\|^2 \leqslant \|\boldsymbol{v} - P_U\boldsymbol{v}\|^2 + \|P_U\boldsymbol{v} - \boldsymbol{u}\|^2$$

6.38 $$= \|(\boldsymbol{v} - P_U\boldsymbol{v}) + (P_U\boldsymbol{v} - \boldsymbol{u})\|^2$$

$$= \|\boldsymbol{v} - \boldsymbol{u}\|^2,$$

其中 6.38 由勾股定理 (6.3) 得到, 之所以能用勾股定理是因为

$v - P_U v \in U^{\perp}$ 并且 $P_U v - u \in U$. 再取平方根即得要证的不等式.

不等式中的等号成立当且仅当 6.37 是等式, 当且仅当 $\|P_U v - u\| = 0$, 当且仅当 $u = P_U v$. ■

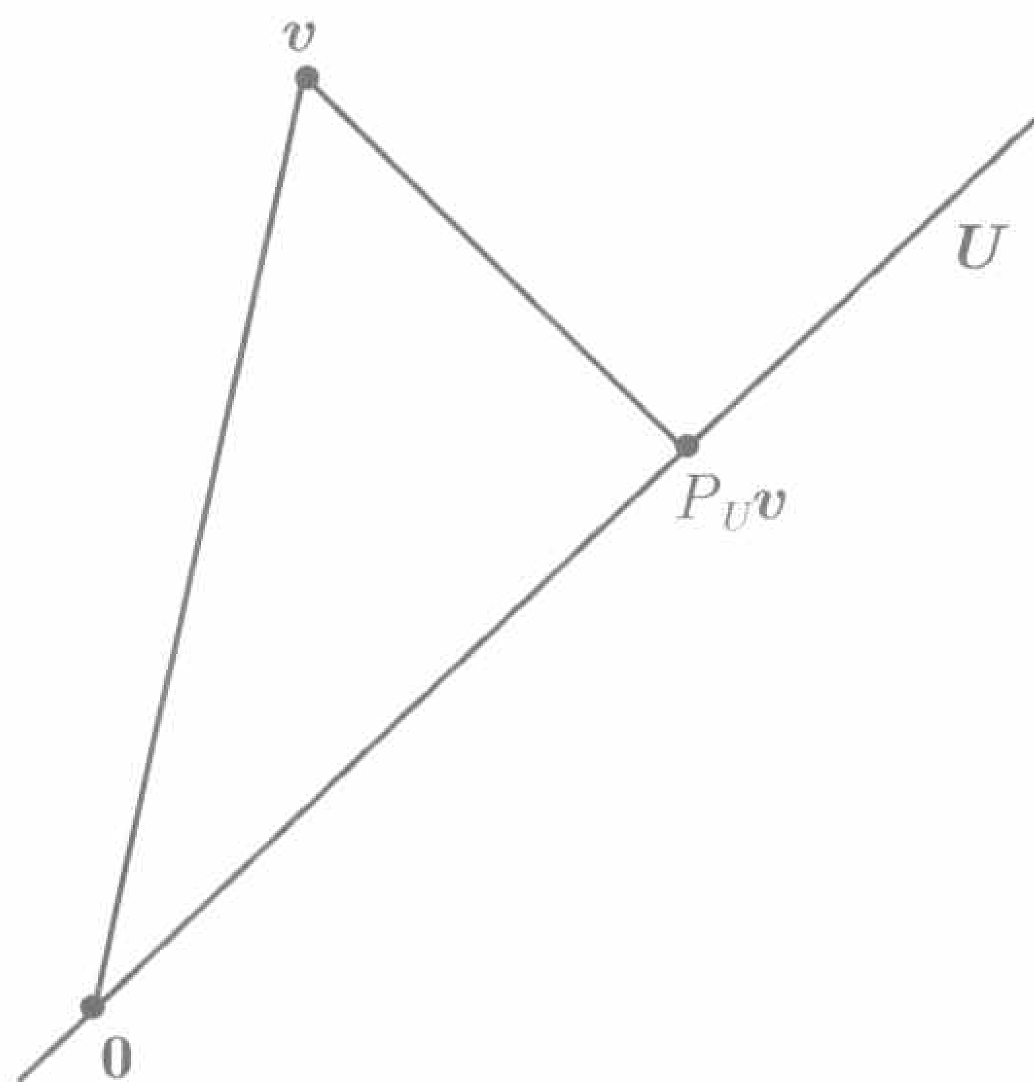


图 6-5　$P_U v$ 是 U 中离 v 最近的点

上一个命题通常与公式 6.35 相结合来计算极小化问题的显式解. 我们来举例说明这一过程. 考虑如下问题: 如何找到一个次数不超过 5 的实系数多项式 u 使其在区间 $[-\pi, \pi]$ 上尽量好地逼近 $\sin x$, 即

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \mathrm{d}x$$

最小. 为解决这一问题, 令 $C[-\pi, \pi]$ 表示由 $[-\pi, \pi]$ 上的连续实值函数组成的实向量空间, 并取内积为:

6.39
$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x) \mathrm{d}x.$$

设 $v \in C[-\pi, \pi]$ 是由 $v(x) = \sin x$ 定义的函数. 令 U 表示由次数不超过 5 的实系数多项式组成的 $C[-\pi, \pi]$ 的子空间. 现在问题可以重述为: 求 $u \in U$ 使得 $\|v - u\|$ 最小.

能够演算积分的计算机在这里是很有用的.

要计算这一逼近问题的解, 首先对 U 的基 $(1, x, x^2, x^3, x^4, x^5)$ 应用格拉姆–施密特过程(利用 6.39 所给出的内积), 得到 U 的规范正交基 $(e_1, e_2, e_3, e_4, e_5, e_6)$. 然后采用 6.39 所给出的内

积, 利用 6.35 计算 $P_U \boldsymbol{v}$ (取 m=6), 可知 $P_U \boldsymbol{v}$ 是如下函数

6.40 $$0.987862x - 0.155271x^3 + 0.00564312x^5,$$

此处我们把精确解中出现的那些 π 换成了一个适当的十进制的近似值.

由 6.36, 上面的多项式是 $\sin x$ 在区间 $[-\pi, \pi]$ 上最佳的 5 次多项式逼近. 要看出这个逼近有多好, 图 6-6 显示了 $\sin x$ 和我们的逼近 6.40 在区间 $[-\pi, \pi]$ 上的图像.

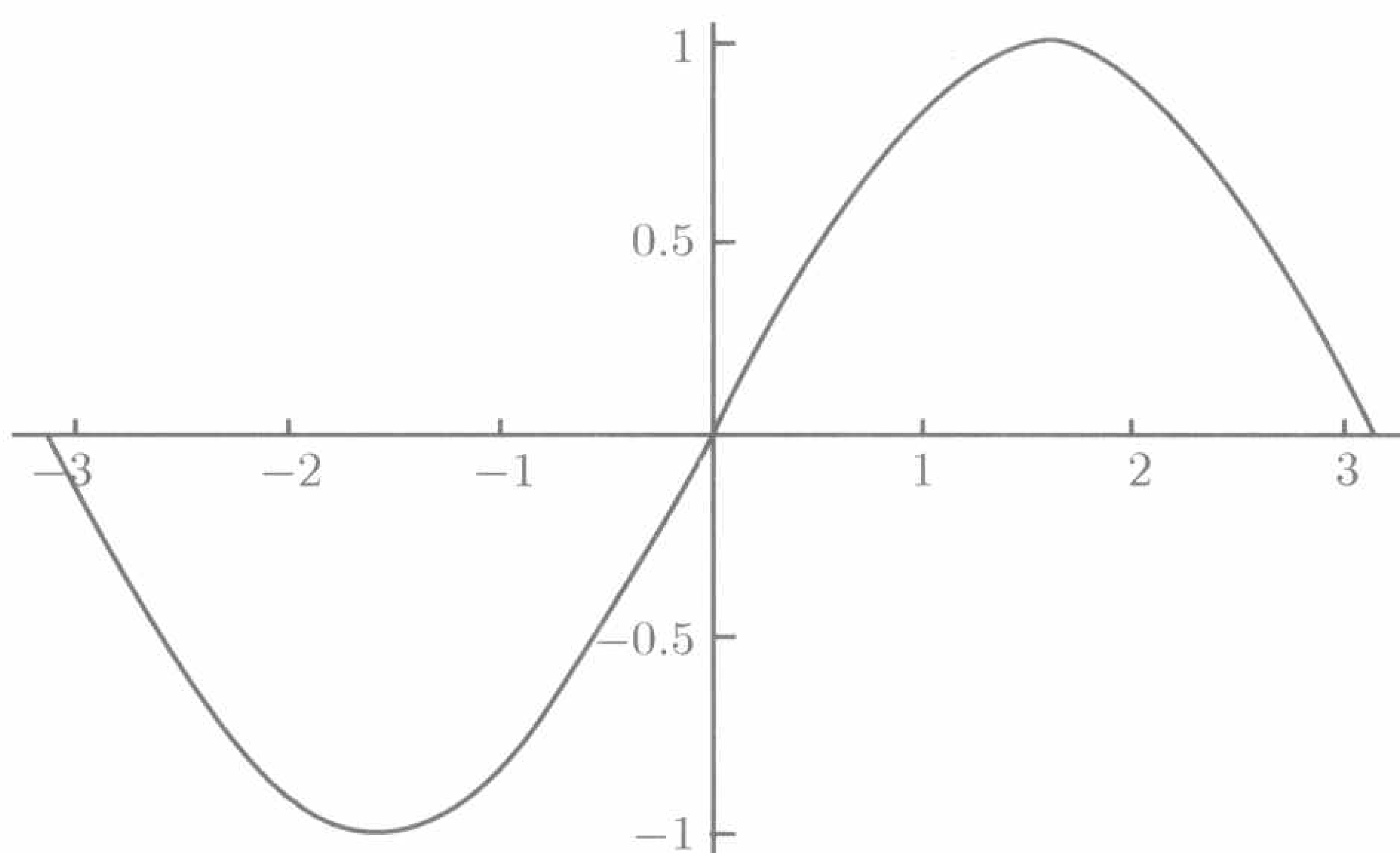


图 6-6 $\sin x$ 及其逼近 6.40 的图像

我们的逼近 6.40 非常精确, 以至于两个图形几乎重合 —— 肉眼只能看到一个图像!

$\sin x$ 的另一个熟知的 5 次多项式逼近是由泰勒多项式给出的

6.41 $$x - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

要看出这个逼近有多好, 图 6-7 显示了 $\sin x$ 和泰勒多项式 6.41 在区间 $[-\pi, \pi]$ 上的图像.

泰勒多项式是在 0 附近对 $\sin x$ 极好的逼近. 但是图 6-7 表明对 $|x| > 2$, Taylor 多项式并不怎么精确, 特别是与 6.40 相比. 例如, 取 $x = 3$, 则 6.40 中的逼近对 $\sin 3$ 估计的误差大概是 0.001, 但是泰勒级数 6.41 对 $\sin 3$ 估计的误差大概是 0.4. 因此对于 $x = 3$, 泰勒级数的误差比 6.40 的误差大几百倍. 线性代数

帮助我们发现了 $\sin x$ 的一个逼近, 这个逼近改进了我们在微积分中所学到的逼近!

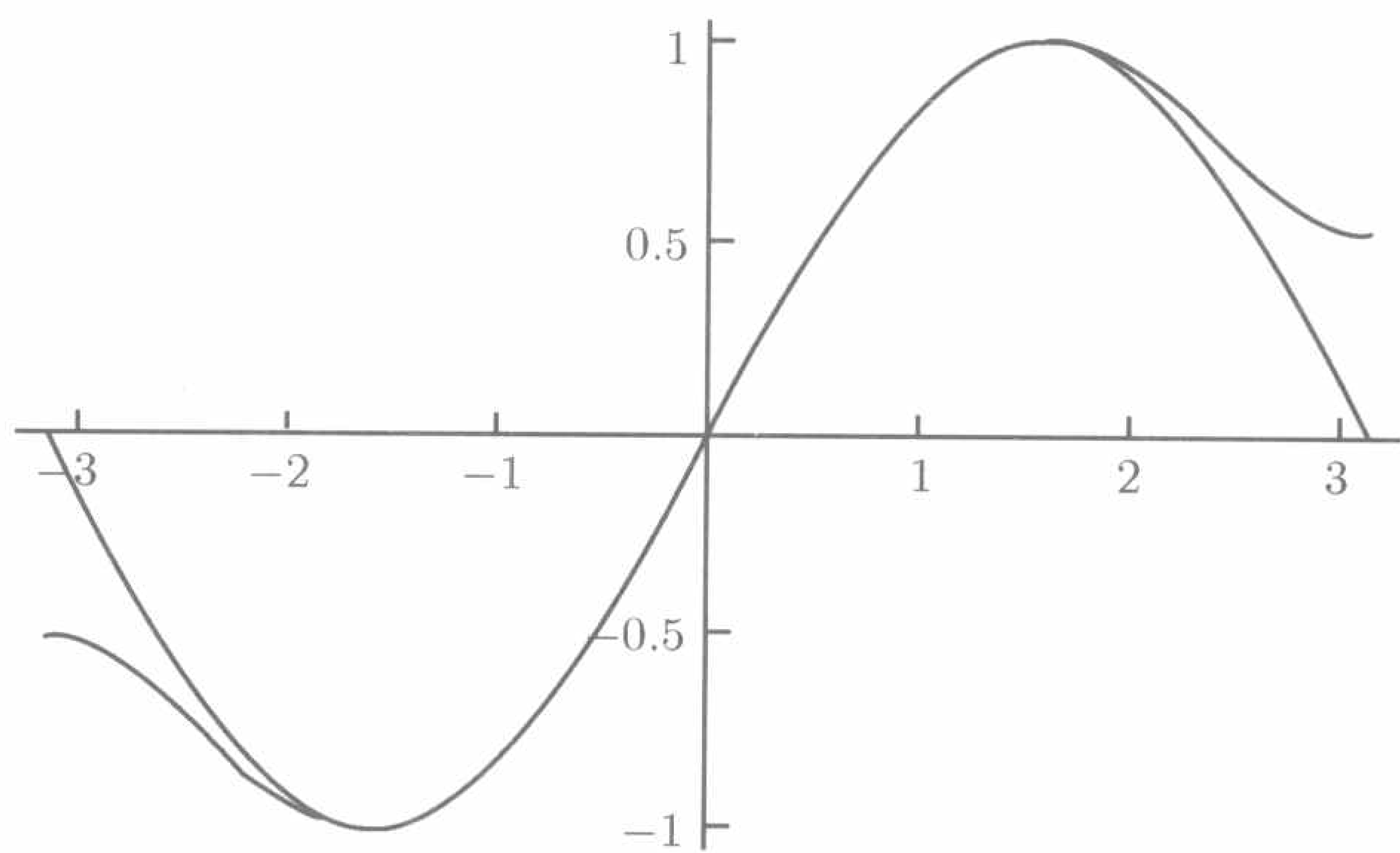


图 6-7 $\sin x$ 和泰勒多项式 6.41 的图像

我们利用 6.35 和 6.36 得到了逼近 6.40. 当 V 等于 $C[-\pi,\pi]$ 时, 标准假设 “V 是有限维向量空间” 不再成立, 因而需要说明在这种情况下这些结果仍然是可用的. 6.29 表明, 若 U 是 V 的子空间, 则

6.42 $$V = U \oplus U^{\perp}.$$

如果我们允许 V 是无限维的, 并且允许 U 是 V 的无限维子空间, 那么在没有额外假设的情况下, 6.42 未必成立.

我们先来回顾一下 6.29 的证明. 这个证明用到了 U 的维数的有限性 (来得到 U 的基), 但无论 V 是否为有限维的, 证明都成立. 其次, 注意到 P_U 的定义和性质 (包括 6.35) 只需要 6.29, 因此只要求 U (而非 V) 是有限维的. 最后, 注意到 6.36 的证明不要求 V 是有限维的. 因此可得出这样的结论: 对 $\boldsymbol{v} \in V$ 和 V 的子空间 U, 无论 V 是否为有限维的, 只要 U 是有限维的, 就可以用上面讨论的过程来寻找向量 $\boldsymbol{u} \in U$ 使得 $\|\boldsymbol{v}-\boldsymbol{u}\|$ 最小. 在上面的例子中, U 确实是有限维的 ($\dim U = 6$), 所以证明都有效.

§6.5 线性泛函与伴随

V 上的**线性泛函** (**linear functional**) 是从 V 到 $\mathbf{F}$ 的线性映射. 例如, 如下定义的函数 $\varphi: \mathbf{F}^3 \to \mathbf{F}$

6.43 $$\varphi(z_1, z_2, z_3) = 2z_1 - 5z_2 + z_3$$

是 $\mathbf{F}^3$ 上的线性泛函. 又如, 考虑内积空间 $\mathcal{P}_6(\mathbf{R})$ (这里的内积是由 [0,1] 上的定积分所诱导的乘法; 参见 6.2). 如下定义的函数 $\varphi: \mathcal{P}_6(\mathbf{R}) \to \mathbf{R}$

6.44 $$\varphi(p) = \int_0^1 p(x)(\cos x)\mathrm{d}x$$

是 $\mathcal{P}_6(\mathbf{R})$ 上的线性泛函.

如果 $\boldsymbol{v} \in V$, 那么把 $\boldsymbol{u}$ 映成 $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ 的映射是 V 上的线性泛函. 下一个结果表明, V 上的每个线性泛函都是这种形式的. 为了举例说明这个定理, 注意到对于 6.43 所定义的线性泛函 φ, 可以取 $\boldsymbol{v} = (2, -5, 1) \in \mathbf{F}^3$. 6.44 所定义的线性泛函 φ 更能显示下面定理的威力, 这是因为对于这个线性泛函, $\boldsymbol{v}$ 没有显然的取法 (函数 $\cos x$ 并不符合条件, 因为它不是 $\mathcal{P}_6(\mathbf{R})$ 中的元素).

6.45 定理: 设 φ 是 V 上的线性泛函, 则存在唯一一个向量 $\boldsymbol{v} \in V$ 使得

$$\varphi(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle, \quad \boldsymbol{u} \in V.$$

证明: 首先证明存在向量 $\boldsymbol{v} \in V$, 使得对每个 $\boldsymbol{u} \in V$ 都有 $\varphi(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle$. 设 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ 是 V 的规范正交基, 则对每个 $\boldsymbol{u} \in V$ 都有

$$\begin{aligned}
\varphi(\boldsymbol{u}) &= \varphi(\langle \boldsymbol{u}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \cdots + \langle \boldsymbol{u}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n) \\
&= \langle \boldsymbol{u}, \boldsymbol{e}_1 \rangle \varphi(\boldsymbol{e}_1) + \cdots + \langle \boldsymbol{u}, \boldsymbol{e}_n \rangle \varphi(\boldsymbol{e}_n) \\
&= \langle \boldsymbol{u}, \overline{\varphi(\boldsymbol{e}_1)} \boldsymbol{e}_1 + \cdots + \overline{\varphi(\boldsymbol{e}_n)} \boldsymbol{e}_n \rangle,
\end{aligned}$$

其中第一个等式由 6.17 得到. 因此, 取 $\boldsymbol{v} = \overline{\varphi(\boldsymbol{e}_1)} \boldsymbol{e}_1 + \cdots + \overline{\varphi(\boldsymbol{e}_n)} \boldsymbol{e}_n$, 则对每个 $\boldsymbol{u} \in V$ 都有 $\varphi(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle$.

现在来证明只有一个向量 $\boldsymbol{v} \in V$ 满足条件. 假设 $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ 使得对每个 $\boldsymbol{u} \in V$ 都有

$$\varphi(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{v}_1 \rangle = \langle \boldsymbol{u}, \boldsymbol{v}_2 \rangle,$$

那么, 对每个 $\boldsymbol{u} \in V$ 都有

$$0 = \langle \boldsymbol{u}, \boldsymbol{v}_1 \rangle - \langle \boldsymbol{u}, \boldsymbol{v}_2 \rangle = \langle \boldsymbol{u}, \boldsymbol{v}_1 - \boldsymbol{v}_2 \rangle.$$

取 $\boldsymbol{u} = \boldsymbol{v}_1 - \boldsymbol{v}_2$ 可得 $\boldsymbol{v}_1 - \boldsymbol{v}_2 = \mathbf{0}$, 即 $\boldsymbol{v}_1 = \boldsymbol{v}_2$, 这就完成了唯一性的证明. ■

除了 V 以外, 我们还需要另外一个有限维内积空间.

我们约定在本章的其余部分, W 是 $\mathbf{F}$ 上的有限维非零内积空间.

设 $T \in \mathcal{L}(V, W)$. T 的**伴随 (adjoint)**, 记为 T^*, 是如下定义的从 W 到 V 的函数. 给定 $\boldsymbol{w} \in W$. 考虑 V 上将 $\boldsymbol{v} \in V$ 映成 $\langle T\boldsymbol{v}, \boldsymbol{w} \rangle$ 的线性泛函. 取 $T^*\boldsymbol{w}$ 是 V 中唯一的那个具有下面性质的向量: 上述线性泛函是通过与 $T^*\boldsymbol{w}$ 做内积所给出的 (6.45 保证了 V 中具有这种性质的向量存在且唯一). 也就是说, $T^*\boldsymbol{w}$ 是 V 中唯一一个满足下面条件的向量

“伴随”这个词在线性代数中还有另一个意思. 在本书中我们不需要这个与逆有关的第二种意思, 只是万一你在别处遇到伴随的第二种含义的话, 要注意, 伴随的这两种意思之间是没有联系的.

$$\langle T\boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{v}, T^*\boldsymbol{w} \rangle, \quad \boldsymbol{v} \in V.$$

我们举例说明如何计算伴随. 定义 $T: \mathbf{R}^3 \to \mathbf{R}^2$ 为

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

因此 T^* 是从 $\mathbf{R}^2$ 到 $\mathbf{R}^3$ 的函数. 要计算 T^*, 固定一个点 $(y_1, y_2) \in \mathbf{R}^2$, 则对所有 $(x_1, x_2, x_3) \in \mathbf{R}^3$ 都有

$$\begin{aligned}
\langle (x_1, x_2, x_3), T^*(y_1, y_2) \rangle &= \langle T(x_1, x_2, x_3), (y_1, y_2) \rangle \\
&= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2) \rangle \\
&= x_2 y_1 + 3x_3 y_1 + 2x_1 y_2 \\
&= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1) \rangle.
\end{aligned}$$

由此得

$$T^*(y_1, y_2) = (2y_2, y_1, 3y_1).$$

注意到在上面的例子中, T^* 不只是从 $\mathbf{R}^2$ 到 $\mathbf{R}^3$ 的函数, 而且还是线性映射. 这是普遍成立的. 具体地, 若 $T \in \mathcal{L}(V, W)$, 则 $T^* \in \mathcal{L}(W, V)$. 为了证明这一点, 设 $T \in \mathcal{L}(V, W)$. 首先验证加性. 给定 $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W$, 则

伴随将在下一章的重要结果中发挥关键作用.

$$\begin{aligned}\langle T\boldsymbol{v}, \boldsymbol{w}_1 + \boldsymbol{w}_2\rangle &= \langle T\boldsymbol{v}, \boldsymbol{w}_1\rangle + \langle T\boldsymbol{v}, \boldsymbol{w}_2\rangle \\ &= \langle \boldsymbol{v}, T^*\boldsymbol{w}_1\rangle + \langle \boldsymbol{v}, T^*\boldsymbol{w}_2\rangle \\ &= \langle \boldsymbol{v}, T^*\boldsymbol{w}_1 + T^*\boldsymbol{w}_2\rangle,\end{aligned}$$

这说明 $T^*\boldsymbol{w}_1 + T^*\boldsymbol{w}_2$ 和 $T^*(\boldsymbol{w}_1 + \boldsymbol{w}_2)$ 扮演了同样的角色. 因为满足这一性质的向量是唯一的, 所以必有

$$T^*\boldsymbol{w}_1 + T^*\boldsymbol{w}_2 = T^*(\boldsymbol{w}_1 + \boldsymbol{w}_2).$$

现在来验证 T^* 的齐性. 若 $a \in \mathbf{F}$, 则

$$\begin{aligned}\langle T\boldsymbol{v}, a\boldsymbol{w}\rangle &= \bar{a}\langle T\boldsymbol{v}, \boldsymbol{w}\rangle \\ &= \bar{a}\langle \boldsymbol{v}, T^*\boldsymbol{w}\rangle \\ &= \langle \boldsymbol{v}, aT^*\boldsymbol{w}\rangle,\end{aligned}$$

这说明 $aT^*\boldsymbol{w}$ 和 $T^*(a\boldsymbol{w})$ 扮演了同样的角色. 因为满足这一性质的向量是唯一的, 所以必有

$$aT^*\boldsymbol{w} = T^*(a\boldsymbol{w}).$$

因此 T^* 是线性映射.

你应该验证, 函数 $T \mapsto T^*$ 具有如下性质:

加性 (additivity)

对所有 $S, T \in \mathcal{L}(V, W)$ 都有 $(S+T)^* = S^* + T^*$;

共轭齐性 (conjugate homogeneity)

对所有 $a \in \mathbf{F}$, $T \in \mathcal{L}(V, W)$ 都有 $(aT)^* = \bar{a}T^*$;

伴随的伴随 (adjoint of adjoint)

对所有 $T \in \mathcal{L}(V, W)$ 都有 $(T^*)^* = T$;

恒等算子 (identity)

$I^* = I$, 其中 I 是 V 上的恒等算子;

乘积 (products)

对所有 $T \in \mathcal{L}(V,W)$, $S \in \mathcal{L}(W,U)$ 都有 $(ST)^* = T^*S^*$ (这里 U 是 $\mathbf{F}$ 上的内积空间).

下一个结果表明了线性映射及其伴随的零空间和值域之间的关系. 记号 $\Longleftrightarrow$的意思是 "当且仅当"; 这个记号也可以理解成 "等价于".

6.46 命题: 设 $T \in \mathcal{L}(V,W)$, 则

(a) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$;

(b) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$;

(c) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$;

(d) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

证明: 首先证明 (a). 设 $\boldsymbol{w} \in W$, 则

$$\begin{aligned}
\boldsymbol{w} \in \operatorname{null} T^* &\Longleftrightarrow T^*\boldsymbol{w} = \mathbf{0} \\
&\Longleftrightarrow \langle \boldsymbol{v}, T^*\boldsymbol{w} \rangle = 0, \boldsymbol{v} \in V \\
&\Longleftrightarrow \langle T\boldsymbol{v}, \boldsymbol{w} \rangle = 0, \boldsymbol{v} \in V \\
&\Longleftrightarrow \boldsymbol{w} \in (\operatorname{range} T)^\perp.
\end{aligned}$$

因此 $\operatorname{null} T^* = (\operatorname{range} T)^\perp$, 这就证明了 (a).

对 (a) 的两端取正交补即得 (d), 这里用到了 6.33. 最后, 在 (a) 和 (d) 中将 T 换成 T^* 即得 (c) 和 (b). ■

一个 $m \times n$ 矩阵的**共轭转置 (conjugate transpose)** 是通过互换行和列, 然后再对每个元素取复共轭所得到的 $n \times m$ 矩阵. 例如,

若 $\mathbf{F} = \mathbf{R}$, 则矩阵的共轭转置等同于它的**转置 (transpose)**, 即通过互换行和列所得到的矩阵.

$$\begin{bmatrix} 2 & 3+4i & 7 \\ 6 & 5 & 8i \end{bmatrix}$$

的共轭转置是矩阵

$$\begin{bmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{bmatrix}.$$

下一个命题说明了怎样通过 T 的矩阵来计算 T^* 的矩阵. 要注意, 下面的命题只有在处理规范正交基时才能使用, 而对于非规范正交基, T^* 的矩阵不一定等于 T 的矩阵的共轭转置.

6.47 命题: 设 $T \in \mathcal{L}(V, W)$. 如果 $(e_1, \cdots, e_n)$ 是 V 的规范正交基, 并且 $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_m)$ 是 W 的规范正交基, 那么

$$\mathcal{M}(T^*, (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_m), (e_1, \cdots, e_n))$$

是

$$\mathcal{M}(T, (e_1, \cdots, e_n), (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_m)).$$

的共轭转置.

线性映射的伴随与基的选取无关. 这解释了为什么我们要强调线性映射的伴随而不是矩阵的共轭转置.

证明: 设 $(e_1, \cdots, e_n)$ 是 V 的规范正交基, $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_m)$ 是 W 的规范正交基. 我们用 $\mathcal{M}(T)$ 来代替更长的记号 $\mathcal{M}(T, (e_1, \cdots, e_n), (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_m))$; 也把 $\mathcal{M}(T^*, (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_m), (e_1, \cdots, e_n))$ 写成 $\mathcal{M}(T^*)$.

回想一下, 把 Te_k 写成这些 $\boldsymbol{f}_j$ 的线性组合可得 $\mathcal{M}(T)$ 的第 k 列; 在这个线性组合中所用到的标量组成了 $\mathcal{M}(T)$ 的第 k 列. 因为 $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_m)$ 是 W 的规范正交基, 所以我们知道怎样把 Te_k 写成这些 $\boldsymbol{f}_j$ 的线性组合 (参见 6.17):

$$Te_k = \langle Te_k, \boldsymbol{f}_1 \rangle \boldsymbol{f}_1 + \cdots + \langle Te_k, \boldsymbol{f}_m \rangle \boldsymbol{f}_m.$$

因此 $\mathcal{M}(T)$ 中第 j 行第 k 列的元素是 $\langle Te_k, \boldsymbol{f}_j \rangle$. 把 T 替换成 T^*, 再互换诸 e 和诸 $\boldsymbol{f}$ 的角色可知, $\mathcal{M}(T^*)$ 中第 j 行第 k 列的元素是 $\langle T^* \boldsymbol{f}_k, e_j \rangle$, 它等于 $\langle \boldsymbol{f}_k, Te_j \rangle$, 这等于 $\overline{\langle Te_j, \boldsymbol{f}_k \rangle}$, 这又等于 $\mathcal{M}(T)$ 中第 k 行第 j 列元素的复共轭. 也就是说, $\mathcal{M}(T^*)$ 等于 $\mathcal{M}(T)$ 的共轭转置. ■

习　题

1. 证明: 如果 $\boldsymbol{x}, \boldsymbol{y}$ 都是 $\mathbf{R}^2$ 中的非零向量, 那么

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|\|\boldsymbol{y}\| \cos\theta,$$

其中 θ 是 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 之间的夹角 (把 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 都看作始于原点的箭头). **提示**: 画出 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的夹角; 然后利用余弦定律.

2. 设 $\boldsymbol{u}, \boldsymbol{v} \in V$. 证明 $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$ 当且仅当

$$\|\boldsymbol{u}\| \leqslant \|\boldsymbol{u} + a\boldsymbol{v}\|, \quad a \in \mathbf{F}.$$

3. 证明对所有实数 $a_1, \cdots, a_n$ 和 $b_1, \cdots, b_n$ 都有

$$\left(\sum_{j=1}^{n} a_j b_j\right)^2 \leqslant \left(\sum_{j=1}^{n} j {a_j}^2\right)\left(\sum_{j=1}^{n} \frac{{b_j}^2}{j}\right).$$

4. 设 $\boldsymbol{u}, \boldsymbol{v} \in V$ 使得

$$\|\boldsymbol{u}\| = 3, \quad \|\boldsymbol{u} + \boldsymbol{v}\| = 4, \quad \|\boldsymbol{u} - \boldsymbol{v}\| = 6.$$

求 $\|\boldsymbol{v}\|$.

5. 证明或反驳: 存在 $\mathbf{R}^2$ 上的内积, 使得相应的范数定义如下

$$\|(x_1, x_2)\| = |x_1| + |x_2|, \quad (x_1, x_2) \in \mathbf{R}^2.$$

6. 证明: 若 V 是实内积空间, 则

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \frac{\|\boldsymbol{u} + \boldsymbol{v}\|^2 - \|\boldsymbol{u} - \boldsymbol{v}\|^2}{4}, \quad \boldsymbol{u}, \boldsymbol{v} \in V.$$

7. 证明: 若 V 是复内积空间, 则对任意 $\boldsymbol{u}, \boldsymbol{v} \in V$ 都有

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \frac{\|\boldsymbol{u} + \boldsymbol{v}\|^2 - \|\boldsymbol{u} - \boldsymbol{v}\|^2 + \|\boldsymbol{u} + i\boldsymbol{v}\|^2 i - \|\boldsymbol{u} - i\boldsymbol{v}\|^2 i}{4}.$$

8. 向量空间 U 上的范数是一个函数 $\|\ \|: U \to [0,\infty)$, 并且具有以下性质: $\|\boldsymbol{u}\| = 0$ 当且仅当 $\boldsymbol{u} = 0$; 对所有 $\alpha \in \mathbf{F}$ 和 $\boldsymbol{u} \in U$ 都有 $\|\alpha\boldsymbol{u}\| = |\alpha|\,\|\boldsymbol{u}\|$; 对所有 $\boldsymbol{u}, \boldsymbol{v} \in U$ 都有 $\|\boldsymbol{u}+\boldsymbol{v}\| \leqslant \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$. 证明: 满足平行四边形等式的范数都来自内积 (即证明: 如果 $\|\ \|$ 是 U 上的一个满足平行四边形等式的范数, 则有 U 上的内积 $\langle\ \rangle$, 使得对所有 $\boldsymbol{u} \in U$ 都有 $\|\boldsymbol{u}\| = \langle\boldsymbol{u},\boldsymbol{u}\rangle^{1/2}$).

9. 设 n 是正整数. 证明

$$\left(\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \cdots, \frac{\sin nx}{\sqrt{\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \cdots, \frac{\cos nx}{\sqrt{\pi}}\right)$$

是 $C[-\pi,\pi]$ 中的规范正交向量组, 其中 $C[-\pi,\pi]$ 是 $[-\pi,\pi]$ 上的连续实值函数所组成的向量空间, 其上的内积为

$$\langle \boldsymbol{f},\boldsymbol{g}\rangle = \int_{-\pi}^{\pi} \boldsymbol{f}(x)\boldsymbol{g}(x)\mathrm{d}x.$$

这个规范正交组经常用来建立诸如潮汐等周期现象的数学模型.

10. 在 $\mathcal{P}_2(\mathbf{R})$ 上考虑内积

$$\langle \boldsymbol{p},\boldsymbol{q}\rangle = \int_{0}^{1} \boldsymbol{p}(x)\boldsymbol{q}(x)\mathrm{d}x.$$

对基 $(1,x,x^2)$ 应用格拉姆–施密特过程求 $\mathcal{P}_2(\mathbf{R})$ 的一个规范正交基.

11. 对非线性无关的向量组应用格拉姆–施密特过程结果会怎样?

12. 设 V 是实内积空间, 并且 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m)$ 是 V 中的线性无关向量组. 证明 V 中恰好有 2^m 个规范正交向量组 $(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_m)$ 使得

$$\operatorname{span}(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j) = \operatorname{span}(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_j), \quad j \in \{1,\cdots,m\}.$$

13. 设 $(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_m)$ 是 V 中的规范正交向量组, 并且 $\boldsymbol{v} \in V$. 证明

$$\|\boldsymbol{v}\|^2 = |\langle\boldsymbol{v},\boldsymbol{e}_1\rangle|^2 + \cdots + |\langle\boldsymbol{v},\boldsymbol{e}_m\rangle|^2$$

当且仅当 $\boldsymbol{v} \in \operatorname{span}(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_m)$.

14. 求 $\mathcal{P}_2(\mathbf{R})$ 的一个规范正交基 (内积如习题 10), 使得 $\mathcal{P}_2(\mathbf{R})$ 上的微分算子 (将 $\boldsymbol{p}$ 映成 $\boldsymbol{p}'$ 的算子) 关于此基具有上三角矩阵.

15. 设 U 是 V 的子空间. 证明

$$\dim U^{\perp} = \dim V - \dim U.$$

16. 设 U 是 V 的子空间. 证明 $U^{\perp} = \{\mathbf{0}\}$ 当且仅当 $U = V$.

17. 证明: 如果 $P \in \mathcal{L}(V)$, 使得 $P^2 = P$ 并且 P 的零空间中的每个向量都正交于 P 的值域中的向量, 那么 P 是正交投影.

18. 证明: 如果 $P \in \mathcal{L}(V)$, 使得 $P^2 = P$ 并且

$$\|P\boldsymbol{v}\| \leqslant \|\boldsymbol{v}\|, \quad \boldsymbol{v} \in V,$$

那么 P 是正交投影.

19. 设 $T \in \mathcal{L}(V)$, 并且 U 是 V 的子空间. 证明 U 在 T 下是不变的当且仅当 $P_U T P_U = T P_U$.

20. 设 $T \in \mathcal{L}(V)$, 并且 U 是 V 的子空间. 证明 U 和 $U^{\perp}$ 在 T 下都是不变的当且仅当 $P_U T = T P_U$.

21. 在 $\mathbf{R}^4$ 中, 设

$$U = \operatorname{span}((1,1,0,0),(1,1,1,2)).$$

求 $\boldsymbol{u} \in U$ 使得 $\|\boldsymbol{u} - (1,2,3,4)\|$ 最小.

22. 求 $\boldsymbol{p} \in \mathcal{P}_3(\mathbf{R})$, 使得 $\boldsymbol{p}(0) = 0$, $\boldsymbol{p}'(0) = 0$, 并且

$$\int_0^1 |2 + 3x - \boldsymbol{p}(x)|^2 \mathrm{d}x$$

最小.

23. 求 $\boldsymbol{p} \in \mathcal{P}_5(\mathbf{R})$ 使得

$$\int_{-\pi}^{\pi} |\sin x - \boldsymbol{p}(x)|^2 \mathrm{d}x$$

最小. (多项式 6.40 是本习题的一个极好的近似解, 但是这里要找的是包含 π 的幂的精确解. 可以使用能进行符号积分的电脑.)

24. 求多项式 $\boldsymbol{q} \in \mathcal{P}_2(\mathbf{R})$ 使得

$$\boldsymbol{p}\left(\frac{1}{2}\right) = \int_0^1 \boldsymbol{p}(x)\boldsymbol{q}(x)\mathrm{d}x, \quad \boldsymbol{p} \in \mathcal{P}_2(\mathbf{R}).$$

25. 求多项式 $\boldsymbol{q} \in \mathcal{P}_2(\mathbf{R})$ 使得

$$\int_0^1 \boldsymbol{p}(x)(\cos \pi x)\mathrm{d}x = \int_0^1 \boldsymbol{p}(x)\boldsymbol{q}(x)\mathrm{d}x, \quad \boldsymbol{p} \in \mathcal{P}_2(\mathbf{R}).$$

26. 取定向量 $\boldsymbol{v} \in V$, 把 $T \in \mathcal{L}(V, \mathbf{F})$ 定义成 $T\boldsymbol{u} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle$. 对 $a \in \mathbf{F}$, 求 T^*a.

27. 设 n 为正整数. 把 $T \in \mathcal{L}(\mathbf{F}^n)$ 定义为

$$T(z_1, \cdots, z_n) = (0, z_1, \cdots, z_{n-1}).$$

求 $T^*(z_1, \cdots, z_n)$.

28. 设 $T \in \mathcal{L}(V)$, 并且 $\lambda \in \mathbf{F}$. 证明 λ 是 T 的本征值当且仅当 $\bar{\lambda}$ 是 T^* 的本征值.

29. 设 $T \in \mathcal{L}(V)$, 并且 U 是 V 的子空间. 证明 U 在 T 下是不变的当且仅当 $U^{\perp}$ 在 T^* 下是不变的.

30. 设 $T \in \mathcal{L}(V, W)$. 证明:
    (a) T 是单的当且仅当 T^* 是满的;
    (b) T 是满的当且仅当 T^* 是单的.

31. 证明对每个 $T \in \mathcal{L}(V, W)$ 都有

$$\dim \operatorname{null} T^* = \dim \operatorname{null} T + \dim W - \dim V,$$

并且

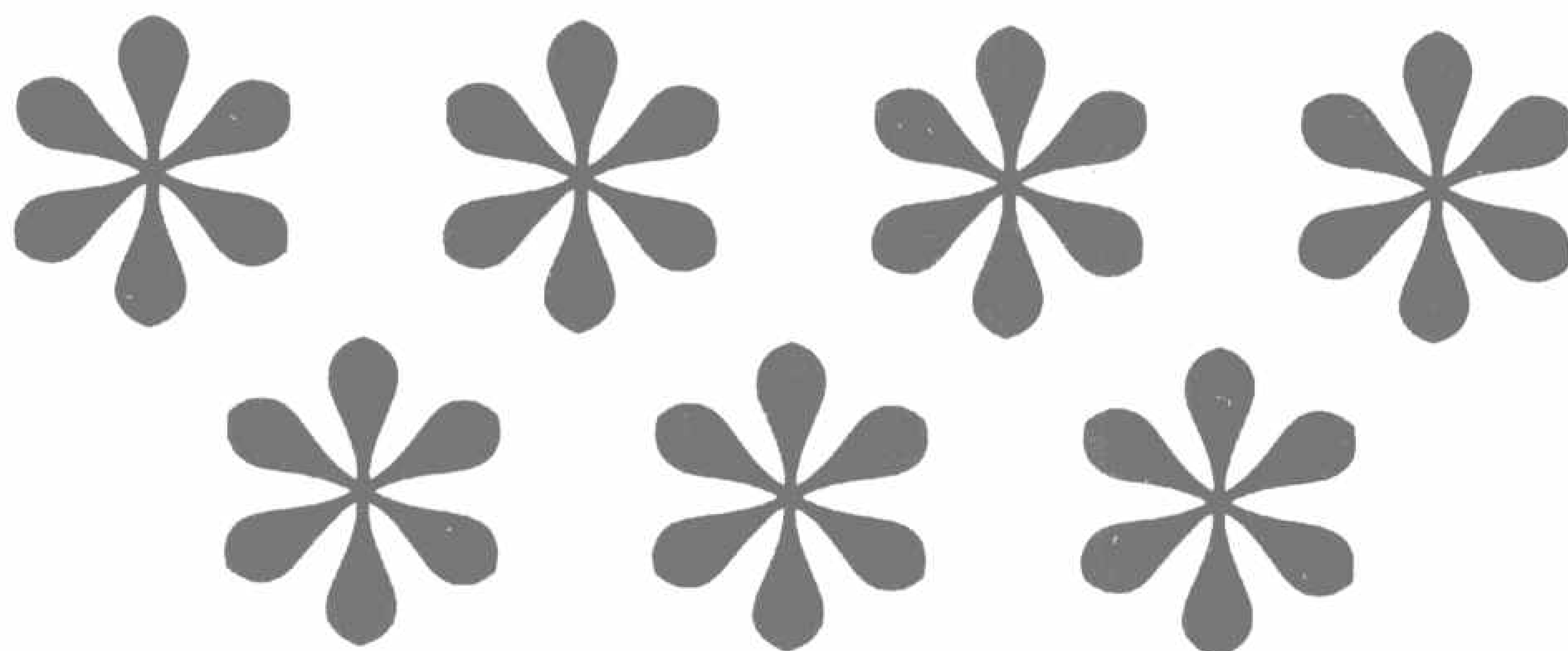
$$\dim \operatorname{range} T^* = \dim \operatorname{range} T.$$

32. 设 $\boldsymbol{A}$ 为 $m \times n$ 实矩阵. 证明: $\boldsymbol{A}$ 的所有列 (在 $\mathbf{R}^m$ 中) 所张成的子空间的维数等于 $\boldsymbol{A}$ 的所有行 (在 $\mathbf{R}^n$ 中) 所张成的子空间的维数.

第 7 章　内积空间上的算子

我们现在要讨论内积空间上的算子，这方面的研究成果在内积空间理论中最为深刻. 我们将利用伴随的性质详细描述内积空间上的几类重要算子.

$\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$.
V 是 $\mathbf{F}$ 上的有限维非零内积空间.



§7.1 自伴算子与正规算子

有的数学家采用术语埃尔米特 (**Hermitian**)代替自伴的 (为了纪念法国数学家埃尔米特 (Charles Hermite), 他在 1873 年首次证明了 e 不是任何整系数多项式的根).

算子 $T \in \mathcal{L}(V)$ 称为**自伴的** (**self-adjoint**), 如果 $T = T^*$. 例如, 若 T 是 $\mathbf{F}^2$ 上的算子, 它 (关于标准基) 的矩阵是

$$\begin{bmatrix} 2 & b \\ 3 & 7 \end{bmatrix},$$

则 T 是自伴的当且仅当 $b = 3$ (这是因为, $\mathcal{M}(T) = \mathcal{M}(T^*)$ 当且仅当 $b = 3$; 回想一下, $\mathcal{M}(T^*)$ 是 $\mathcal{M}(T)$ 的共轭转置 —— 参见 6.47).

你应该验证, 两个自伴算子的和是自伴的, 一个实数和一个自伴算子的乘积是自伴的.

要记住这样的类比 (尤其当 $\mathbf{F} = \mathbf{C}$ 时): 伴随在 $\mathcal{L}(V)$ 上所起的作用犹如复共轭在 $\mathbf{C}$ 上所起的作用. 复数 z 是实的当且仅当 $z = \bar{z}$; 因而, 自伴算子 ($T = T^*$) 可与实数类比. 我们将看到, 这种类比也反映在自伴算子的某些重要性质上, 先来看本征值.

若 $\mathbf{F} = \mathbf{R}$, 则由定义可知每个本征值都是实的, 故这个命题仅当 $\mathbf{F} = \mathbf{C}$ 时才有意思.

7.1 命题: 自伴算子的本征值都是实的.

证明: 设 T 是 V 上的自伴算子, λ 是 T 的本征值, 并且 $\boldsymbol{v}$ 是 V 中的非零向量使得 $T\boldsymbol{v} = \lambda\boldsymbol{v}$, 则

$$\begin{aligned}
\lambda\|\boldsymbol{v}\|^2 &= \langle \lambda\boldsymbol{v}, \boldsymbol{v} \rangle \\
&= \langle T\boldsymbol{v}, \boldsymbol{v} \rangle \\
&= \langle \boldsymbol{v}, T\boldsymbol{v} \rangle \\
&= \langle \boldsymbol{v}, \lambda\boldsymbol{v} \rangle \\
&= \bar{\lambda}\|\boldsymbol{v}\|^2.
\end{aligned}$$

于是, $\lambda = \bar{\lambda}$, 即 λ 是实的. ■

下一个命题对实内积空间不成立. 例如, 考虑算子 $T \in \mathcal{L}(\mathbf{R}^2)$, 它是绕原点的反时针 90° 旋转, 因此 $T(x, y) = (-y, x)$. 虽然 T 不是 0, 但是对每个 $\boldsymbol{v} \in \mathbf{R}^2$, $T\boldsymbol{v}$ 显然正交于 $\boldsymbol{v}$.

7.2 命题: 若 V 是复内积空间, T 是 V 上的算子, 使得对所有 $\boldsymbol{v} \in V$ 都有

$$\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0,$$

则 $T = 0$.

证明: 设 V 是复内积空间, $T \in \mathcal{L}(V)$, 则对所有 $\boldsymbol{u}, \boldsymbol{w} \in V$, 通过计算可得

$$\begin{aligned}\langle T\boldsymbol{u}, \boldsymbol{w}\rangle =& \frac{\langle T(\boldsymbol{u}+\boldsymbol{w}), \boldsymbol{u}+\boldsymbol{w}\rangle - \langle T(\boldsymbol{u}-\boldsymbol{w}), \boldsymbol{u}-\boldsymbol{w}\rangle}{4} \\ &+ \frac{\langle T(\boldsymbol{u}+i\boldsymbol{w}), \boldsymbol{u}+i\boldsymbol{w}\rangle - \langle T(\boldsymbol{u}-i\boldsymbol{w}), \boldsymbol{u}-i\boldsymbol{w}\rangle}{4} i.\end{aligned}$$

注意到右端的每一项都具有 $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle$ 的形式. 如果对所有 $\boldsymbol{v} \in V$ 都有 $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0$, 则由上式可知, 对所有 $\boldsymbol{u}, \boldsymbol{w} \in V$ 都有 $\langle T\boldsymbol{u}, \boldsymbol{w}\rangle = 0$, 从而 $T = 0$ (取 $\boldsymbol{w} = T\boldsymbol{u}$). ■

通过考虑实内积空间上的非自伴算子可知, 下面的推论对实内积空间不成立.

7.3 推论: 设 V 是复内积空间, $T \in \mathcal{L}(V)$, 则 T 是自伴的当且仅当对每个 $\boldsymbol{v} \in V$ 都有

$$\langle T\boldsymbol{v}, \boldsymbol{v}\rangle \in \mathbf{R}.$$

这个推论提供了自伴算子与实数具有相似性质的另一个例子.

证明: 设 $\boldsymbol{v} \in V$, 则

$$\begin{aligned}\langle T\boldsymbol{v}, \boldsymbol{v}\rangle - \overline{\langle T\boldsymbol{v}, \boldsymbol{v}\rangle} &= \langle T\boldsymbol{v}, \boldsymbol{v}\rangle - \langle \boldsymbol{v}, T\boldsymbol{v}\rangle \\ &= \langle T\boldsymbol{v}, \boldsymbol{v}\rangle - \langle T^*\boldsymbol{v}, \boldsymbol{v}\rangle \\ &= \langle (T - T^*)\boldsymbol{v}, \boldsymbol{v}\rangle.\end{aligned}$$

若对每个 $\boldsymbol{v} \in V$ 都有 $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle \in \mathbf{R}$, 则上式左端等于 0, 故对每个 $\boldsymbol{v} \in V$ 都有 $\langle (T - T^*)\boldsymbol{v}, \boldsymbol{v}\rangle = 0$, 从而 $T - T^* = 0$ (由 7.2), 因此 T 是自伴的.

反之, 若 T 是自伴的, 则上式的右端等于 0, 故对每个 $\boldsymbol{v} \in V$ 都有 $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = \overline{\langle T\boldsymbol{v}, \boldsymbol{v}\rangle}$, 从而对每个 $\boldsymbol{v} \in V$ 都有 $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle \in \mathbf{R}$. ■

在实内积空间 V 上, 对于非零算子 T 可能会出现这样的情况: 对所有 $\boldsymbol{v} \in V$ 都有 $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0$. 然而, 下一个命题表明, 对于非零自伴算子就不会出现这种情况.

7.4 命题: 若 T 是 V 上的自伴算子使得对所有 $\boldsymbol{v} \in V$ 都有

$$\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0,$$

则 $T = 0$.

证明: 我们已经对复内积空间证明了这一结论 (没有 T 自伴的假设, 参见 7.2). 因此, 可设 V 是实内积空间, 并且 T 是 V 上的自伴算子, 从而对 $\boldsymbol{u}, \boldsymbol{w} \in V$, 有

$$\begin{aligned}\langle T\boldsymbol{w}, \boldsymbol{u}\rangle &= \langle \boldsymbol{w}, T\boldsymbol{u}\rangle \\ &= \langle T\boldsymbol{u}, \boldsymbol{w}\rangle,\end{aligned}$$

由此直接计算可得

7.5 $$\langle T\boldsymbol{u}, \boldsymbol{w}\rangle = \frac{\langle T(\boldsymbol{u}+\boldsymbol{w}), \boldsymbol{u}+\boldsymbol{w}\rangle - \langle T(\boldsymbol{u}-\boldsymbol{w}), \boldsymbol{u}-\boldsymbol{w}\rangle}{4}.$$

若对所有 $\boldsymbol{v} \in V$ 都有 $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0$, 则由 7.5 可知, 对所有 $\boldsymbol{u}, \boldsymbol{w} \in V$ 都有 $\langle T\boldsymbol{u}, \boldsymbol{w}\rangle = 0$, 从而 $T = 0$ (取 $\boldsymbol{w} = T\boldsymbol{u}$). ■

内积空间上的算子称为**正规的 (normal)**, 如果它和它的伴随交换; 也就是说, $T \in \mathcal{L}(V)$ 是正规的当且仅当

$$TT^* = T^*T.$$

自伴算子显然是正规的. 为了举出一个非自伴的正规算子的例子, 考虑 $\mathbf{F}^2$ 上 (关于标准基) 具有如下矩阵的算子,

$$\begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix}.$$

这个算子显然不是自伴的, 但是容易证明它是正规的 (你应该证一下).

我们马上就会看到正规算子值得特别关注的原因. 下一个命题给出了正规算子的一个简单刻划.

7.6 命题： 算子 $T \in \mathcal{L}(V)$ 是正规的当且仅当对所有 $v \in V$ 都有

$$\|Tv\| = \|T^*v\|.$$

注意，这个命题意味着每个正规算子 T 都满足 $\operatorname{null} T = \operatorname{null} T^*$.

证明： 设 $T \in \mathcal{L}(V)$. 我们将同时证明这个结果的两个方面. 注意到

$$\begin{aligned} T\text{是正规的} &\iff T^*T - TT^* = 0 \\ &\iff \langle (T^*T - TT^*)v, v\rangle = 0, \quad v \in V \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle, \quad v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2, \quad v \in V, \end{aligned}$$

其中第二个等价性由 7.4 得到 (注意到算子 $T^*T - TT^*$ 是自伴的). 第一个条件和最后一个条件的等价性给出了要证明的结果. ■

把下一个推论同前一章的习题 28 对比一下. 那个习题说明, 算子的伴随的所有本征值 (作为集合) 等于该算子所有本征值的复共轭; 但是该习题中没有关于本征向量的论断, 这是因为算子与其伴随可以有不同的本征向量. 然而, 由下一个推论可知, 正规算子与其伴随有相同的本征向量.

7.7 推论： 设 $T \in \mathcal{L}(V)$ 是正规的. 若 $v \in V$ 是 T 的相应于本征值 $\lambda \in \mathbf{F}$ 的本征向量, 则 v 也是 T^* 的相应于本征值 $\bar{\lambda}$ 的本征向量.

证明： 设 $v \in V$ 是 T 的相应于本征值 λ 的本征向量, 则 $(T - \lambda I)v = \mathbf{0}$. 你应该验证, 因为 T 是正规的, 所以 $T - \lambda I$ 也是正规的. 利用 7.6 可得

$$0 = \|(T - \lambda I)v\| = \|(T - \lambda I)^*v\| = \|(T^* - \bar{\lambda}I)v\|,$$

因此 v 是 T^* 的相应于本征值 $\bar{\lambda}$ 的本征向量. ■

因为自伴算子是正规的, 所以下一个结果也适用于自伴算子.

7.8 推论： 如果 $T \in \mathcal{L}(V)$ 是正规的, 那么 T 的相应于不同本征值的本征向量是正交的.

证明： 设 $T \in \mathcal{L}(V)$ 是正规的, α, β 是 T 的不同本征值, $\boldsymbol{u}, \boldsymbol{v}$ 分别是相应的本征向量, 则 $T\boldsymbol{u} = \alpha\boldsymbol{u}$, $T\boldsymbol{v} = \beta\boldsymbol{v}$. 由 7.7 可得 $T^*\boldsymbol{v} = \bar{\beta}\boldsymbol{v}$. 于是

$$\begin{aligned}(\alpha - \beta)\langle \boldsymbol{u}, \boldsymbol{v}\rangle &= \langle \alpha\boldsymbol{u}, \boldsymbol{v}\rangle - \langle \boldsymbol{u}, \bar{\beta}\boldsymbol{v}\rangle \\ &= \langle T\boldsymbol{u}, \boldsymbol{v}\rangle - \langle \boldsymbol{u}, T^*\boldsymbol{v}\rangle \\ &= 0\end{aligned}$$

因为 $\alpha \neq \beta$, 所以由上式可得 $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$. 故 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 是正交的.

■

§7.2　谱　定　理

回想一下, 对角矩阵是对角线之外的元素都是 0 的方阵; V 上的算子关于某个基有对角矩阵当且仅当 V 有一个由该算子的本征向量组成的基 (参见 5.21).

关于某个**规范正交基**具有对角矩阵的算子是 V 上最好的算子, 它们恰好是具有如下性质的算子 $T \in \mathcal{L}(V)$: V 有一个由 T 的本征向量组成的规范正交基. 本节的目的是证明谱定理. 谱定理表明, 具有上述性质的算子恰好是复正规算子或实自伴算子. 谱定理可能是研究内积空间上算子的最有用的工具.

因为谱定理的结论依赖于 $\mathbf{F}$, 所以我们把谱定理分成两部分, 分别叫做复谱定理和实谱定理. 同线性代数中的多数情形一样, 研究复向量空间要比研究实向量空间容易, 因此我们先给出复谱定理.

作为复谱定理的一个例证, 考虑正规算子 $T \in \mathcal{L}(\mathbf{C}^2)$, 它 (关于标准基) 的矩阵是

$$\begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix}.$$

你应该验证,

$$\left(\frac{(i,1)}{\sqrt{2}},\frac{(-i,1)}{\sqrt{2}}\right)$$

是 $\mathbf{C}^2$ 的由 T 的本征向量组成的规范正交基, 并且 T 关于此基的矩阵是对角矩阵

$$\begin{bmatrix} 2+3i & 0 \\ 0 & 2-3i \end{bmatrix}.$$

7.9 复谱定理 (Complex Spectral Theorem): 设 V 是复内积空间, 并且 $T \in \mathcal{L}(V)$, 则 V 有一个由 T 的本征向量组成的规范正交基当且仅当 T 是正规的.

因为自伴算子都是正规的, 所以复谱定理意味着有限维复内积空间上的自伴算子关于某个规范正交基有对角矩阵.

证明: 首先假设 V 有一个由 T 的本征向量组成的规范正交基, 则 T 关于此基有对角矩阵. T^* (关于同一个基) 的矩阵显然是 T 的矩阵的共轭转置, 故 T^* 也有对角矩阵. 任意两个对角矩阵都交换, 故 T 和 T^* 交换, 从而 T 是正规的.

为了证明另一个方面, 现在假设 T 是正规的, 则 V 有规范正交基 $(e_1,\cdots,e_n)$ 使得 T 关于此基有上三角矩阵 (由 6.28). 于是,

$$\mathcal{M}(T,(e_1,\cdots,e_n)) = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{bmatrix}. \tag{7.10}$$

我们将证明这个矩阵实际上是对角矩阵, 即 $(e_1,\cdots,e_n)$ 是 V 的由 T 的本征向量组成的规范正交基.

由上面的矩阵可得,

$$\|Te_1\|^2 = |a_{1,1}|^2$$

并且

$$\|T^*e_1\|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2.$$

因为 T 是正规的, 所以 $\|Te_1\| = \|T^*e_1\|$ (参见 7.6). 于是由上面的两个等式可知, 7.10 中矩阵的第一行除了第一个元素 $a_{1,1}$ 之外都等于 0.

现在由 7.10 可得

$$\|Te_2\|^2 = |a_{2,2}|^2$$

(因为如上一段所证, $a_{1,2}=0$) 并且

$$\|T^*e_2\|^2 = |a_{2,2}|^2 + |a_{2,3}|^2 + \cdots + |a_{2,n}|^2.$$

因为 T 是正规的, 所以 $\|Te_2\| = \|T^*e_2\|$. 于是由上面的两个等式知, 7.10 中矩阵的第二行除了对角线元素 $a_{2,2}$ 之外都等于 0.

如此继续下去可知, 7.10 中矩阵的非对角线元素都等于 0.

■

为证明实谱定理, 我们需要两个引理. 你可能会猜到下一个引理, 甚至可能会通过考虑实系数的二次多项式给出其证明. 具体地, 假设 $\alpha, \beta \in \mathbf{R}$ 使得 $\alpha^2 < 4\beta$, 并设 x 是一个实数, 则

配方法可以用来导出二次求根公式.

$$\begin{aligned} x^2 + \alpha x + \beta &= \left(x + \frac{\alpha}{2}\right)^2 + \left(\beta - \frac{\alpha^2}{4}\right) \\ &> 0. \end{aligned}$$

特别地, $x^2 + \alpha x + \beta$ 是一个可逆的实数 (非 0 的一种迂回的说法). 用一个自伴算子代替实数 x (回想一下实数和自伴算子之间的类比) 可以得出下面的引理.

7.11 引理: 设 $T \in \mathcal{L}(V)$ 是自伴的. 若 $\alpha, \beta \in \mathbf{R}$ 使得 $\alpha^2 < 4\beta$, 则

$$T^2 + \alpha T + \beta I$$

是可逆的.

证明: 设 $\alpha, \beta \in \mathbf{R}$ 使得 $\alpha^2 < 4\beta$, 并设 $\boldsymbol{v}$ 是 V 中的非零向量, 则

$$\begin{aligned} \langle (T^2 + \alpha T + \beta I)\boldsymbol{v}, \boldsymbol{v} \rangle &= \langle T^2\boldsymbol{v}, \boldsymbol{v} \rangle + \alpha \langle T\boldsymbol{v}, \boldsymbol{v} \rangle + \beta \langle \boldsymbol{v}, \boldsymbol{v} \rangle \\ &= \langle T\boldsymbol{v}, T\boldsymbol{v} \rangle + \alpha \langle T\boldsymbol{v}, \boldsymbol{v} \rangle + \beta \|\boldsymbol{v}\|^2 \\ &\geqslant \|T\boldsymbol{v}\|^2 - |\alpha| \|T\boldsymbol{v}\| \|\boldsymbol{v}\| + \beta \|\boldsymbol{v}\|^2 \\ &= \left(\|T\boldsymbol{v}\| - \frac{|\alpha| \|\boldsymbol{v}\|}{2}\right)^2 + \left(\beta - \frac{\alpha^2}{4}\right) \|\boldsymbol{v}\|^2 \\ &> 0, \end{aligned}$$

其中第一个不等式成立是由于柯西–施瓦茨不等式 (6.6). 由上面

最后的那个不等式可得 $(T^2+\alpha T+\beta I)\boldsymbol{v}\neq\mathbf{0}$, 从而 $T^2+\alpha T+\beta I$ 是单的, 因此是可逆的 (参见 3.21). ■

我们已经证明了, 有限维复向量空间上的算子, 无论自伴与否, 都有本征值 (参见 5.10), 因此下一个引理只对实内积空间是新的.

7.12 引理: 设 $T\in\mathcal{L}(V)$ 是自伴的, 则 T 有本征值.

证明: 如前所述, 可设 V 是实内积空间, 并设 $n=\dim V$. 取 $\boldsymbol{v}\in V$ 使得 $\boldsymbol{v}\neq\mathbf{0}$. 因为 V 是 n 维的, 所以下面的 $n+1$ 个向量

$$(\boldsymbol{v}, T\boldsymbol{v}, T^2\boldsymbol{v},\cdots,T^n\boldsymbol{v})$$

不可能是线性无关的. 于是, 有不全为 0 的实数 $a_0,\cdots,a_n$ 使得

这里模仿了 T 有1维或2维不变子空间的证明 (参见 5.24).

$$\mathbf{0}=a_0\boldsymbol{v}+a_1T\boldsymbol{v}+\cdots+a_nT^n\boldsymbol{v}.$$

以这些 a 为系数作一个多项式, 并将此多项式分解成 (参见 4.14)

$$\begin{aligned}&a_0+a_1x+\cdots+a_nx^n\\ &=c(x^2+\alpha_1x+\beta_1)\cdots(x^2+\alpha_Mx+\beta_M)(x-\lambda_1)\cdots(x-\lambda_m),\end{aligned}$$

其中 c 是非零实数, 每个 α_j, β_j, λ_j 都是实的, $\alpha_j^2<4\beta_j$, $m+M\geqslant 1$, 并且上面的等式对所有实的 x 都成立. 因此, 我们有

$$\begin{aligned}0&=a_0\boldsymbol{v}+a_1T\boldsymbol{v}+\cdots+a_nT^n\boldsymbol{v}\\ &=(a_0I+a_1T+\cdots+a_nT^n)\boldsymbol{v}\\ &=c(T^2+\alpha_1T+\beta_1I)\cdots(T^2+\alpha_MT+\beta_MI)\\ &\quad\times(T-\lambda_1I)\cdots(T-\lambda_mI)\boldsymbol{v}.\end{aligned}$$

因为 T 是自伴的, 并且 $\alpha_j^2<4\beta_j$ (参见 7.11), 所以每个 $T^2+\alpha_jT+\beta_jI$ 都可逆. 由于 $c\neq0$, 故由上面的等式可得

$$\mathbf{0}=(T-\lambda_1I)\cdots(T-\lambda_mI)\boldsymbol{v}.$$

因此, 至少有一个 j 使得 $T-\lambda_jI$ 不是单的. 也就是说, T 有本征值. ■

作为实谱定理的一个例证, 考虑 $\mathbf{R}^3$ 上的自伴算子 T, 它 (关于标准基) 的矩阵是

$$\begin{bmatrix} 14 & -13 & 8 \\ -13 & 14 & 8 \\ 8 & 8 & -7 \end{bmatrix}.$$

你应该验证

$$\left(\frac{(1,-1,0)}{\sqrt{2}}, \frac{(1,1,1)}{\sqrt{3}}, \frac{(1,1,-2)}{\sqrt{6}}\right)$$

是 $\mathbf{R}^3$ 的由 T 的本征向量组成的规范正交基, 并且 T 关于此基的矩阵是对角矩阵

$$\begin{bmatrix} 27 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & -15 \end{bmatrix}.$$

把复谱定理和实谱定理合并到一起可以得出结论: V 上每个自伴算子关于某个规范正交基都有对角矩阵. 这个结论无论是对 $\mathbf{F} = \mathbf{C}$ 还是对 $\mathbf{F} = \mathbf{R}$ 都成立, 它是谱定理最有用的部分.

7.13 实谱定理 (Real Spectral Theorem): 设 V 是实内积空间, $T \in \mathcal{L}(V)$, 则 V 有一个由 T 的本征向量组成的规范正交基当且仅当 T 是自伴的.

证明: 首先假设 V 有一个由 T 的本征向量组成的规范正交基, 则 T 关于此基有对角矩阵. 这个矩阵等于它的共轭转置. 因此 $T = T^*$, 即 T 是自伴的.

为了证明另一个方面, 现在假设 T 是自伴的. 对 V 的维数用归纳法, 往证 V 有一个由 T 的本征向量组成的规范正交基. 首先注意到, 若 $\dim V = 1$, 则结果显然成立. 现在假设 $\dim V > 1$, 并且假设在维数更小的向量空间上结果成立.

证明思路是, 任取 T 的一个范数为 1 的本征向量 $\boldsymbol{u}$, 然后再添加由 $T|_{\{\boldsymbol{u}\}^\perp}$ 的本征向量组成的 $\{\boldsymbol{u}\}^\perp$ 的规范正交基. 至于细节, 最重要的是验证 $T|_{\{\boldsymbol{u}\}^\perp}$ 是自伴的 (于是我们可以使用归纳法假设).

设 λ 是 T 的任意本征值 (因为 T 是自伴的, 所以由引理 7.2 可知它有本征值), 并设 $\boldsymbol{u} \in V$ 是相应的一个本征向量且 $\|\boldsymbol{u}\| = 1$, 再设 U 表示由 $\boldsymbol{u}$ 的所有标量倍组成的 V 的 1 维子空间. 注意到一个向量 $\boldsymbol{v} \in V$ 含于 $U^\perp$ 当且仅当 $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$.

用一个非零本征向量除以它的范数就可以得到一个范数为 1 的本征向量.

设 $\boldsymbol{v} \in U^\perp$, 则因 T 是自伴的, 故有

$$\langle \boldsymbol{u}, T\boldsymbol{v} \rangle = \langle T\boldsymbol{u}, \boldsymbol{v} \rangle = \langle \lambda\boldsymbol{u}, \boldsymbol{v} \rangle = \lambda\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$$

因此, $T\boldsymbol{v} \in U^\perp$. 于是, 当 $\boldsymbol{v} \in U^\perp$ 时, 也有 $T\boldsymbol{v} \in U^\perp$. 也就是说, $U^\perp$ 在 T 下是不变的, 从而 $S = T|_{U^\perp}$ 定义了一个算子 $S \in \mathcal{L}(U^\perp)$. 若 $\boldsymbol{v}, \boldsymbol{w} \in U^\perp$, 则

$$\langle S\boldsymbol{v}, \boldsymbol{w} \rangle = \langle T\boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{v}, T\boldsymbol{w} \rangle = \langle \boldsymbol{v}, S\boldsymbol{w} \rangle,$$

这就证明了 S 是自伴的 (注意到在上式中间的等式中我们用到了 T 的自伴性). 因而, 由归纳法假设, $U^\perp$ 有一个由 S 的本征向量组成的规范正交基. 显然, S 的每个本征向量都是 T 的本征向量 (因为对每个 $\boldsymbol{v} \in U^\perp$ 都有 $S\boldsymbol{v} = T\boldsymbol{v}$). 于是, 把 $\boldsymbol{u}$ 添加到由 S 的本征向量组成的 $U^\perp$ 的规范正交基就得到了 V 的一个由 T 的本征向量组成的规范正交基. ■

对于自伴的 $T \in \mathcal{L}(V)$ (或更一般地, 当 $\mathbf{F} = \mathbf{C}$ 时, 正规的 $T \in \mathcal{L}(V)$), 下面的推论给出了 V 的最可能的不变子空间直和分解. 在每个 $\operatorname{null}(T - \lambda_j I)$ 上, 算子 T 的作用相当于乘以 λ_j.

7.14 推论: 设 $T \in \mathcal{L}(V)$ 是自伴的 (或当 $\mathbf{F} = \mathbf{C}$ 时, $T \in \mathcal{L}(V)$ 是正规的). 令 $\lambda_1, \cdots, \lambda_m$ 表示 T 的所有互不相同的本征值, 那么

$$V = \operatorname{null}(T - \lambda_1 I) \oplus \cdots \oplus \operatorname{null}(T - \lambda_m I).$$

进一步, $\operatorname{null}(T - \lambda_j I)$ 中的向量正交于此分解中其他子空间中的向量.

证明: 由谱定理 (7.9 和 7.13) 可知, V 有一个由 T 的本征向量组成的基. 现在由 5.21 即得想要的分解.

由 7.8 可得关于正交性的结论. ■

§7.3 实内积空间上的正规算子

复谱定理 (7.9) 完全描述了复内积空间上的正规算子. 本节将完全描述实内积空间上的正规算子. 在此过程中, 我们会遇到一个命题 (7.18) 和一种技巧 (分块对角矩阵), 它们对于实内积空间和复内积空间都是有用的.

先来描述 2 维实内积空间上非自伴的正规算子.

7.15 引理: 设 V 是 2 维实内积空间, $T \in \mathcal{L}(V)$, 则下列等价:

(a) T 是正规的, 但不是自伴的;

(b) T 关于 V 的每个规范正交基的矩阵都具有如下形式

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad b \neq 0;$$

(c) T 关于 V 的某个规范正交基的矩阵具有如下形式

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad b > 0.$$

证明: 首先假设 (a) 成立, 则 T 是正规的, 但不是自伴的. 设 (e_1, e_2) 是 V 的规范正交基, 并设

7.16 $$\mathcal{M}(T, (e_1, e_2)) = \begin{bmatrix} a & c \\ b & d \end{bmatrix},$$

则 $\|Te_1\|^2 = a^2 + b^2$, $\|T^*e_1\|^2 = a^2 + c^2$. 因为 T 是正规的, 所以 $\|Te_1\| = \|T^*e_1\|$ (参见 7.6); 于是, 由这些等式可得 $b^2 = c^2$, 故 $c = b$ 或 $c = -b$. 由 7.16 中的矩阵可知 $c \neq b$, 因为否则 T 就是自伴的. 于是, $c = -b$, 故

7.17 $$\mathcal{M}(T, (e_1, e_2)) = \begin{bmatrix} a & -b \\ b & d \end{bmatrix}.$$

当然, T^* 的矩阵是上面这个矩阵的转置. 用矩阵乘法计算 TT^*

和 T^*T 的矩阵 (现在就动手算一下). 因为 T 是正规的, 这两个矩阵一定相等. 比较你算出来的这两个矩阵右上角的元素, 就会发现 $bd = ab$. 现在由 7.17 中的矩阵可知 $b \neq 0$, 因为否则 T 就是自伴的. 于是, $d = a$, 故 (a) 蕴含 (b).

现在假设 (b) 成立, 往证 (c) 成立. 任取 V 的一个规范正交基 $(\boldsymbol{e}_1, \boldsymbol{e}_2)$, 则 T 关于此基的矩阵具有 (b) 所给出的形式, 且 $b \neq 0$. 若 $b > 0$, 则 (c) 成立, 从而证明了 (b) 蕴含 (c). 如果 $b < 0$, 那么你应该验证, T 关于规范正交基 $(\boldsymbol{e}_1, -\boldsymbol{e}_2)$ 的矩阵等于 $\left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$, 其中 $-b > 0$; 于是, 对于这种情形, 仍有 (b) 蕴含 (c).

现在假设 (c) 成立, 则 T 关于某个规范正交基的矩阵具有 (c) 所给出的形式, 且 $b > 0$. 显然 T 的矩阵与其转置不相等 (因为 $b \neq 0$), 因此 T 不是自伴. 现在用矩阵乘法可以验证 TT^* 的矩阵和 T^*T 的矩阵是相等的. 因此, $TT^* = T^*T$, 故 T 正规. 于是 (c) 蕴含 (a). ■

我们将引进一种记号用来把一个矩阵写成由一些小矩阵组成的矩阵, 例如, 考虑矩阵

$$\boldsymbol{D} = \begin{bmatrix} 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 \\ 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 3 & 3 & 3 \end{bmatrix}.$$

我们可将此矩阵写成如下形式

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix},$$

其中

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix},$$

并且$\boldsymbol{0}$表示由 0 组成的 3×2 矩阵.

为了更好地理解一个矩阵, 我们经常把它看成是由一些更小的矩阵组成的. 我们将在下一个命题和以后的各章中使用这种技巧.

下一个结果将在刻画实内积空间上的正规算子时发挥关键作用.

即使没有正规性, 仍有一个更简单的结果成立: 若 $T \in \mathcal{L}(V)$, 并且 U 在 T 下是不变的, 则 $U^\perp$ 在 T^* 下是不变的; 参见第 6 章习题 29.

7.18 命题: 设 $T \in \mathcal{L}(V)$ 是正规的, 并且 U 是 V 的在 T 下的不变子空间, 则

(a) $U^\perp$ 在 T 下是不变的;

(b) U 在 T^* 下是不变的;

(c) $(T|_U)^* = (T^*)|_U$;

(d) $T|_U$ 是 U 上的正规算子;

(e) $T|_{U^\perp}$ 是 $U^\perp$ 上的正规算子.

证明: 先证明 (a). 设 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 是 U 的规范正交基, 将其扩充成 V 的规范正交基, $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m, \boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ (由 6.25 知这是可能的). 因为 U 在 T 下是不变的, 所以每个 $T\boldsymbol{e}_j$ 都是 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 的线性组合. 于是, T 关于基 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m, \boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ 的矩阵具有如下形式

$$\mathcal{M}(T) = \begin{array}{c} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_m \\ \boldsymbol{f}_1 \\ \vdots \\ \boldsymbol{f}_n \end{array} \overset{\begin{array}{cccccc} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_m & \boldsymbol{f}_1 & \cdots & \boldsymbol{f}_n \end{array}}{\begin{bmatrix} \boldsymbol{A} & & \boldsymbol{B} \\ & & \\ \boldsymbol{0} & & \boldsymbol{C} \end{bmatrix}};$$

其中 $\boldsymbol{A}$ 表示一个 $m \times m$ 矩阵, $\boldsymbol{0}$ 表示由 0 组成的 $n \times m$ 矩阵, $\boldsymbol{B}$ 表示一个 $m \times n$ 矩阵, $\boldsymbol{C}$ 表示一个 $n \times n$ 矩阵, 并且为了方便, 所选取的基已列在矩阵的顶端和左侧.

对每个 $j \in \{1, \cdots, m\}$, $\|T\boldsymbol{e}_j\|^2$ 等于 $\boldsymbol{A}$ 的第 j 列元素绝对值的平方和 (参见 6.17). 因此

7.19 $$\sum_{j=1}^{m} \|T\boldsymbol{e}_j\|^2 = \boldsymbol{A} \text{ 中元素绝对值的平方和}.$$

对每个 $j \in \{1, \cdots, m\}$, $\|T^*\boldsymbol{e}_j\|^2$ 都等于 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的第 j 列元素绝对值的平方和. 因此

7.20 $$\sum_{j=1}^{m}\|T^*\boldsymbol{e}_j\|^2 = \boldsymbol{A} \text{ 和 } \boldsymbol{B} \text{ 中元素绝对值的平方和}.$$

因为 T 是正规的, 所以对每个 j 都有 $\|T\boldsymbol{e}_j\| = \|T^*\boldsymbol{e}_j\|$ (参见 7.6). 于是

$$\sum_{j=1}^{m}\|T\boldsymbol{e}_j\|^2 = \sum_{j=1}^{m}\|T^*\boldsymbol{e}_j\|^2.$$

根据 7.19 和 7.20, 此式意味着 $\boldsymbol{B}$ 中元素绝对值的平方和必须等于 0. 也就是说, $\boldsymbol{B}$ 一定是由 0 组成的矩阵. 于是

7.21 $$\mathcal{M}(T) = \begin{array}{c} \\ \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_m \\ \boldsymbol{f}_1 \\ \vdots \\ \boldsymbol{f}_n \end{array} \begin{array}{c} \boldsymbol{e}_1 \ \cdots \ \boldsymbol{e}_m \ \boldsymbol{f}_1 \ \cdots \ \boldsymbol{f}_n \\ \left[\begin{array}{cc} \boldsymbol{A} & 0 \\ 0 & \boldsymbol{C} \end{array}\right] \end{array}.$$

这说明, 对每个 k, $T\boldsymbol{f}_k$ 都含于 $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ 的张成. 因为 $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ 是 $U^\perp$ 的基, 所以这意味着, 当 $\boldsymbol{v} \in U^\perp$ 时必有 $T\boldsymbol{v} \in U^\perp$. 也就是说, $U^\perp$ 在 T 下是不变的, 这就证明了 (a).

为证 (b), 注意到 $\mathcal{M}(T^*)$ 左下角有一块全是 0 (因为 $\mathcal{M}(T)$ 右上角有一块全是 0). 也就是说, 每个 $T^*\boldsymbol{e}_j$ 都可以写成 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m)$ 的线性组合. 于是, U 在 T^* 下是不变的, 这就证明了 (b).

为证 (c), 令 $S = T|_U$, 并固定 $\boldsymbol{v} \in U$, 则对所有 $\boldsymbol{u} \in U$ 都有

$$\begin{aligned}\langle S\boldsymbol{u}, \boldsymbol{v}\rangle &= \langle T\boldsymbol{u}, \boldsymbol{v}\rangle \\ &= \langle \boldsymbol{u}, T^*\boldsymbol{v}\rangle.\end{aligned}$$

因为 $T^*\boldsymbol{v} \in U$ (由 (b)), 所以上式表明 $S^*\boldsymbol{v} = T^*\boldsymbol{v}$. 也就是说, $(T|_U)^* = (T^*)|_U$, 这就证明了 (c).

为证 (d), 注意到 T 和 T^* 可交换 (因为 T 是正规的), 而且 $(T|_U)^* = (T^*)|_U$ (由 (c)). 于是, $T|_U$ 与其伴随可交换, 故正规, 这就证明了 (d).

为证 (e), 注意到在 (d) 中我们证明了 T 到任何不变子空间上的限制都是正规的. 然而, $U^{\perp}$ 在 T 下是不变的 (由 (a)), 故 $T|_{U^{\perp}}$ 是正规的. ■

在命题 7.18 的证明中, 关键步骤是证明 $\mathcal{M}(T)$ 是一个适当的分块对角矩阵; 参见 7.21.

在 7.18 的证明过程中, 我们把一个矩阵看成是由一些更小的矩阵组成的. 现在我们需要更多地运用这种思想. **分块对角矩阵 (block diagonal matrix)** 就是如下形式的方阵

$$\begin{bmatrix} \boldsymbol{A}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 都是方阵, 它们位于对角线上, 而矩阵的其他元素都等于 0. 例如, 矩阵

7.22
$$\boldsymbol{A} = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 2 & -3 & 0 & 0 \\ 0 & 3 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 7 & 1 \end{bmatrix}$$

是分块对角矩阵, 即有

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & & 0 \\ & \boldsymbol{A}_2 & \\ 0 & & \boldsymbol{A}_3 \end{bmatrix},$$

其中

7.23
$$\boldsymbol{A}_1 = \begin{bmatrix} 4 \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix}, \quad \boldsymbol{A}_3 = \begin{bmatrix} 1 & -7 \\ 7 & 1 \end{bmatrix}.$$

如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是如下形式的分块对角矩阵

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{B}_m \end{bmatrix},$$

其中 $\boldsymbol{A}_j$ 和 $\boldsymbol{B}_j$ 大小相同, $j=1,\cdots,m$, 那么你应该验证, $\boldsymbol{AB}$ 是如下形式的分块对角矩阵,

7.24
$$\begin{bmatrix} \boldsymbol{A}_1\boldsymbol{B}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{A}_m\boldsymbol{B}_m \end{bmatrix},$$

也就是说, 两个 (具有相同分块的) 分块对角矩阵相乘, 像对角矩阵相乘一样, 只要把对角线上相应的块相乘即可.

对角矩阵是一种特殊的分块对角矩阵, 其中每个块的大小都是 1×1. 另一个极端的情形是, 每个方阵都是分块对角矩阵, 因为我们可以把第一个 (也是唯一的) 块取成整个矩阵. 因而, 说算子关于某个基有分块对角矩阵等于什么也没说, 除非我们知道这些块的大小, 块越小, 算子就越好 (含糊地说, 就是算子的矩阵有更多的 0). 最好的情形是存在规范正交基使得算子关于此基有对角矩阵. 我们已经证明, 在复内积空间上, 这种情形恰好对正规算子才会出现 (参见 7.9), 而在实内积空间上恰好对自伴算子才会出现 (参见 7.13).

注意到如果算子 T 关于某个基有分块对角矩阵, 则此矩阵对角线上的任何 1×1 块都一定是 T 的本征值.

下一个结果表明, 实内积空间上的正规算子的矩阵接近于对角矩阵 —— 具体地, 关于某个规范正交基, 我们得到了分块对角矩阵, 而且每个块的大小最多是 2×2. 我们不能指望分得更好, 因为实内积空间上确实有一些正规算子关于任何基都没有对角矩阵. 例如, (你应该验证) 由 $T(x,y)=(-y,x)$ 定义的算子 $T\in\mathcal{L}(\mathbf{R}^2)$ 是正规的, 但是没有本征值; 因此这个特殊的 T 关于 $\mathbf{R}^2$ 的任何基甚至都不会有上三角矩阵.

注意到 7.22 中的矩阵就是下面的定理所保证的那种类型的矩阵. 特别地, 7.22 中每个块 (参见 7.23) 的大小都不超过 2×2, 而且每个 2×2 块都具有所要求的形式 (左上方的元素等于右下方的元素, 并且左下方元素都是正的, 而右上方元素都等于左下方元素的相反数).

7.25 定理: 假设 V 是实内积空间, 并且 $T\in\mathcal{L}(V)$, 则 T 是正规的当且仅当 V 有规范正交基使得 T 关于此基有分块对角矩阵, 而且每个块都是 1×1 矩阵或如下形式的 2×2 矩阵

7.26
$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad b>0.$$

证明： 为了证明容易的一方面，首先假设 V 有规范正交基使得 T 的矩阵是分块对角矩阵，并且每个块都是 1×1 矩阵或形如 7.26 的 2×2 矩阵. 你应该验证 (用公式 7.24 求两个分块对角矩阵之积)，关于这个基，T 的矩阵和 T^* 的矩阵 (它是 T 的矩阵的共轭转置) 交换. 于是，T 和 T^* 交换，故 T 是正规的.

为证另一方面，现在假设 T 是正规的. 对 V 的维数用归纳法. 首先注意到，若 $\dim V=1$ (平凡地)，或 $\dim V=2$ (若 T 自伴，则用实谱定理 7.13; 否则用 7.15)，则结果显然成立.

现在假设 $\dim V>2$，并且结果在维数更小的向量空间上成立. 如果 V 有在 T 下不变的 1 维子空间，则令 U 是其中之一 (也就是说，若 T 有非零本征向量，则令 U 是由这个本征向量张成的子空间). 如果没有这样的子空间，则令 U 是在 T 下不变的 2 维子空间 (由 5.24 可知，1 维或 2 维的不变子空间总是存在的).

在 1 维实向量空间中，恰好有两个范数为 1 的向量.

如果 $\dim U=1$，则在 U 中取一个范数为 1 的向量；这个向量就是 U 的规范正交基，$T|_U$ 的矩阵当然是 1×1 矩阵. 如果 $\dim U=2$，则 $T|_U$ 是正规的 (由 7.18)，但不是自伴的 (否则，$T|_U$ 以及 T 就会有非零本征向量；参见 7.12)，于是可取 U 的规范正交基使得 $T|_U$ 关于此基的矩阵具有 7.26 的形式 (参见 7.15).

现在 $U^\perp$ 在 T 下是不变的，并且 $T|_{U^\perp}$ 是 $U^\perp$ 上的正规算子 (参见 7.18). 因而由归纳法假设，$U^\perp$ 有规范正交基使得 $T|_{U^\perp}$ 关于此基的矩阵具有我们想要的形式. 把这个基添加到 U 的基就得到 V 的规范正交基，并且 T 关于此基的矩阵具有我们想要的形式. ■

§7.4 正 算 子

许多数学家也采用**半正定算子** (positive semidefinite operator) 这个术语，意思与正算子相同.

对于算子 $T\in\mathcal{L}(V)$，如果 T 是自伴的，并且对所有 $\boldsymbol{v}\in V$ 都有

$$\langle T\boldsymbol{v},\boldsymbol{v}\rangle\geqslant 0,$$

则称 T 为**正的** (**positive**). 注意到，若 V 是复向量空间，则 T 自伴的条件可以去掉 (由 7.3).

你应该验证, 正交投影都是正的. 至于其他例子, 看一下 7.11 的证明, 其中我们证明了: 若 $T \in \mathcal{L}(V)$ 是自伴的, 并且 $\alpha, \beta \in \mathbf{R}$ 使得 $\alpha^2 < 4\beta$, 则 $T^2 + \alpha T + \beta I$ 是正的.

算子 S 称为算子 T 的**平方根 (square root)**, 如果 $S^2 = T$. 例如, 若 $T \in \mathcal{L}(\mathbf{F}^3)$ 是由 $T(z_1, z_2, z_3) = (z_3, 0, 0)$ 定义的算子, $S \in \mathcal{L}(\mathbf{F}^3)$ 是由 $S(z_1, z_2, z_3) = (z_2, z_3, 0)$ 定义的算子, 则 S 是 T 的平方根.

下面的定理是关于正算子的主要结果. 注意到正算子的刻画与 $\mathbf{C}$ 中非负数的刻画是对应的. 具体地, 复数 z 非负当且仅当它有非负平方根, 这对应于下面的条件 (c). 此外, z 非负当且仅当它有实的平方根, 这对应于下面的条件 (d). 最后, z 非负当且仅当有复数 w 使得 $z = \bar{w}w$, 这对应于下面的条件 (e).

在某种意义上, 正算子相应于 $[0, \infty)$ 中的数, 因此在术语上更应该称之为非负的而不是正的. 然而, 算子理论学家始终称之为正算子, 因此我们将沿袭这个传统.

7.27 定理: 设 $T \in \mathcal{L}(V)$, 则下列等价:

(a) T 是正的;
(b) T 是自伴的, 且 T 的所有本征值都非负;
(c) T 有正平方根;
(d) T 有自伴的平方根;
(e) 有算子 $S \in \mathcal{L}(V)$ 使得 $T = S^*S$.

证明: 我们将证明 (a)⇒ (b)⇒ (c)⇒ (d)⇒ (e)⇒ (a).

首先假设 (a) 成立, 因此 T 是正的. 显然 T 是自伴的 (根据正算子的定义). 为证 (b) 中的其他结果, 设 λ 是 T 的本征值, $\boldsymbol{v}$ 是 T 的相应于 λ 的非零本征向量, 则

$$\begin{aligned} 0 &\leqslant \langle T\boldsymbol{v}, \boldsymbol{v}\rangle \\ &= \langle \lambda\boldsymbol{v}, \boldsymbol{v}\rangle \\ &= \lambda\langle \boldsymbol{v}, \boldsymbol{v}\rangle, \end{aligned}$$

于是 λ 是非负数. 因此 (b) 成立.

现在假设 (b) 成立, 因此 T 是自伴的, 并且 T 的所有本征值都是非负的. 由谱定理 (7.9 和 7.13), V 有一个由 T 的本征向量组成的规范正交基 $(e_1, \cdots, e_n)$. 设相应的本征值为 $\lambda_1, \cdots, \lambda_n$, 则每个 λ_j 都是非负数. 定义 $S \in \mathcal{L}(V)$ 如下

$$Se_j = \sqrt{\lambda_j} e_j, \quad j = 1, \cdots, n,$$

那么你应该验证, S 是正算子. 进一步, 对每个 j 都有 $S^2 e_j = \lambda_j e_j = T e_j$, 由此可得 $S^2 = T$. 于是, S 是 T 的正平方根, 故 (c) 成立.

显然 (c) 蕴含 (d) (因为, 由定义知正算子都是自伴的).

现在假设 (d) 成立, 即有 V 上的自伴算子 S 使得 $T = S^2$. 那么 $T = S^*S$ (因为 $S^* = S$), 故 (e) 成立.

最后, 假设 (e) 成立. 设 $S \in \mathcal{L}(V)$ 使得 $T = S^*S$, 则 $T^* = (S^*S)^* = S^*(S^*)^* = S^*S = T$, 故 T 是自伴的. 为证 (a) 成立, 注意到对每个 $v \in V$ 都有

$$\begin{aligned} \langle Tv, v \rangle &= \langle S^*Sv, v \rangle \\ &= \langle Sv, Sv \rangle \\ &\geqslant 0. \end{aligned}$$

于是, T 是正的. ■

每个非负数都有唯一一个非负平方根. 下一个命题表明正算子也具有类似的性质. 根据这个命题, 我们可以采用记号 $\sqrt{T}$来表示正算子 T 唯一的正平方根, 正如 $\sqrt{\lambda}$ 表示非负数 λ 唯一的非负平方根.

正算子可能有无穷多个平方根 (但是其中只有一个是正的). 例如, 若 $\dim V > 1$, 则 V 上的恒等算子就有无穷多个平方根.

7.28 命题: *V 上每个正算子都有唯一一个正平方根.*

证明: 设 $T \in \mathcal{L}(V)$ 是正的, 并设 $\lambda_1, \cdots, \lambda_m$ 表示 T 的所有互不相同的本征值; 因为 T 是正的, 所以这些数都非负 (由 7.27). 因为 T 是自伴的, 所以由 7.14 可得

7.29 $$V = \operatorname{null}(T - \lambda_1 I) \oplus \cdots \oplus \operatorname{null}(T - \lambda_m I).$$

现在设 $S \in \mathcal{L}(V)$ 是 T 的正平方根, 并设 α 是 S 的本征值. 若 $\boldsymbol{v} \in \operatorname{null}(S-\alpha I)$, 则 $S\boldsymbol{v}=\alpha\boldsymbol{v}$, 从而

7.30 $$T\boldsymbol{v}=S^2\boldsymbol{v}=\alpha^2\boldsymbol{v},$$

故 $\boldsymbol{v} \in \operatorname{null}(T-\alpha^2 I)$. 于是 α^2 是 T 的本征值, 由此可知 α^2 必定等于某个 λ_j. 也就是说, 有某个 j 使得 $\alpha=\sqrt{\lambda_j}$. 进一步, 由 7.30 可得

7.31 $$\operatorname{null}(S-\sqrt{\lambda_j}I) \subset \operatorname{null}(T-\lambda_j I).$$

我们在上一段证明了, S 的本征值只能是 $\sqrt{\lambda_1},\cdots,\sqrt{\lambda_m}$. 因为 S 是自伴的, 所以由 7.14 可得

7.32 $$V=\operatorname{null}(S-\sqrt{\lambda_1}I)\oplus\cdots\oplus\operatorname{null}(S-\sqrt{\lambda_m}I).$$

现在, 由 7.29, 7.32 以及 7.31 可知, 对每个 j 都有

$$\operatorname{null}(S-\sqrt{\lambda_j}I)=\operatorname{null}(T-\lambda_j I).$$

也就是说, 在 $\operatorname{null}(T-\lambda_j I)$ 上, 算子 S 的作用相当于乘以 $\sqrt{\lambda_j}$. 于是, T 的正平方根 S 是由 T 唯一确定的. ■

§7.5 等距同构

算子 $S \in \mathcal{L}(V)$ 称为**等距同构 (isometry)**, 如果对所有 $\boldsymbol{v} \in V$ 都有

$$\|S\boldsymbol{v}\|=\|\boldsymbol{v}\|.$$

也就是说, 算子是等距同构当且仅当它保范数. 例如, 当 $\lambda \in \mathbf{F}$ 满足 $|\lambda|=1$ 时, λI 是等距同构. 更一般地, 设 $\lambda_1,\cdots,\lambda_n$ 都是绝对值为 1 的标量, $(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)$ 是 V 的规范正交基, $S \in \mathcal{L}(V)$ 满足 $S(\boldsymbol{e}_j)=\lambda_j\boldsymbol{e}_j$. 对于 $\boldsymbol{v} \in V$, 由 6.17 可知

在希腊语中, isos 的意思是相等; metron 的意思是度量. 因此 isometry (等距同构) 的字面意思就是度量相等.

7.33 $$\boldsymbol{v}=\langle\boldsymbol{v},\boldsymbol{e}_1\rangle\boldsymbol{e}_1+\cdots+\langle\boldsymbol{v},\boldsymbol{e}_n\rangle\boldsymbol{e}_n,$$

而且

7.34 $$\|\boldsymbol{v}\|^2=|\langle\boldsymbol{v},\boldsymbol{e}_1\rangle|^2+\cdots+|\langle\boldsymbol{v},\boldsymbol{e}_n\rangle|^2.$$

把 S 作用到 7.33 的两端可得

$$\begin{aligned} Sv &= \langle v, e_1\rangle Se_1 + \cdots + \langle v, e_n\rangle Se_n \\ &= \lambda_1\langle v, e_1\rangle e_1 + \cdots + \lambda_n\langle v, e_n\rangle e_n. \end{aligned}$$

根据 $|\lambda_j| = 1$, 上面最后的等式表明

7.35 $$\|Sv\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2.$$

比较 7.34 和 7.35 可得 $\|v\| = \|Sv\|$. 也就是说, S 是等距同构.

实内积空间上的等距同构通常称为**正交 (orthogonal) 算子**. 复内积空间上的等距同构通常称为**酉 (unitary) 算子**. 我们将采用等距同构这个术语, 因此我们的结果对于实内积空间和复内积空间都可用.

又如, 设 $\theta \in \mathbf{R}$, 则 $\mathbf{R}^2$ 上 (以原点为中心) 反时针旋转 θ 角度所确定的算子是等距同构 (你应该找出这个算子关于 $\mathbf{R}^2$ 的标准基的矩阵).

若 $S \in \mathcal{L}(V)$ 是等距同构, 则 S 是单的 (因为, 若 $Sv = \mathbf{0}$, 则 $\|v\| = \|Sv\| = 0$, 故 $v = \mathbf{0}$). 于是每个等距同构都可逆 (由 3.21).

下一个定理提供了等距同构的若干等价条件. 这些等价条件有一些重要解释. 特别地, (a) 和 (b) 的等价性表明等距同构保内积. 由于 (a) 蕴含 (d), 从而若 S 是等距同构而 $(e_1, \cdots, e_n)$ 是 V 的规范正交基, 则 S (关于此基) 的矩阵的列是规范正交的; 又因 (e) 蕴含 (a), 故逆命题也成立. 因为 (a) 等价于条件 (i) 和 (j), 所以在上一句话中也可用 “行” 代替 “列”.

7.36 定理: 设 $S \in \mathcal{L}(V)$, 则下列等价:

(a) S 是等距同构;
(b) 对所有 $u, v \in V$, 都有 $\langle Su, Sv\rangle = \langle u, v\rangle$;
(c) $S^*S = I$;
(d) 若 $(e_1, \cdots, e_n)$ 是 V 中的规范正交向量组, 则 $(Se_1, \cdots, Se_n)$ 是规范正交的;
(e) V 有规范正交基 $(e_1, \cdots, e_n)$ 使得 $(Se_1, \cdots, Se_n)$ 是规范正交的;
(f) S^* 是等距同构;
(g) 对所有 $u, v \in V$, 都有 $\langle S^*u, S^*v\rangle = \langle u, v\rangle$;

(h) $SS^* = I$;

(i) 若 $(e_1, \cdots, e_n)$ 是 V 中的规范正交向量组, 则 $(S^*e_1, \cdots, S^*e_n)$ 是规范正交的;

(j) V 有规范正交基 $(e_1, \cdots, e_n)$ 使得 $(S^*e_1, \cdots, S^*e_n)$ 是规范正交的.

证明: 首先假设 (a) 成立. 如果 V 是实内积空间, 那么对每个 $\boldsymbol{u}, \boldsymbol{v} \in V$ 都有

$$\begin{aligned}\langle S\boldsymbol{u}, S\boldsymbol{v}\rangle &= (\|S\boldsymbol{u} + S\boldsymbol{v}\|^2 - \|S\boldsymbol{u} - S\boldsymbol{v}\|^2)/4 \\ &= (\|S(\boldsymbol{u} + \boldsymbol{v})\|^2 - \|S(\boldsymbol{u} - \boldsymbol{v})\|^2)/4 \\ &= (\|\boldsymbol{u} + \boldsymbol{v}\|^2 - \|\boldsymbol{u} - \boldsymbol{v}\|^2)/4 \\ &= \langle \boldsymbol{u}, \boldsymbol{v}\rangle,\end{aligned}$$

其中第一个等式可以由第 6 章习题 6 推出, 第二个等式可以由 S 的线性推出, 第三个等式成立是因为 S 是等距同构, 而最后的等式也是由第 6 章习题 6 推出的. 若 V 是复内积空间, 则用第 6 章的习题 7 代替习题 6 可以得到同样的结论. 无论哪种情况, 都有 (a) 蕴含 (b).

现在假设 (b) 成立, 则对任意 $\boldsymbol{u}, \boldsymbol{v} \in V$ 都有

$$\begin{aligned}\langle (S^*S - I)\boldsymbol{u}, \boldsymbol{v}\rangle &= \langle S\boldsymbol{u}, S\boldsymbol{v}\rangle - \langle \boldsymbol{u}, \boldsymbol{v}\rangle \\ &= 0.\end{aligned}$$

取 $\boldsymbol{v} = (S^*S - I)\boldsymbol{u}$ 可得, $S^*S - I = 0$. 因此 $S^*S = I$, 这就证明了 (b) 蕴含 (c).

现在假设 (c) 成立. 设 $(e_1, \cdots, e_n)$ 是 V 中的规范正交向量组, 则

$$\begin{aligned}\langle Se_j, Se_k\rangle &= \langle S^*Se_j, e_k\rangle \\ &= \langle e_j, e_k\rangle.\end{aligned}$$

因此, $(Se_1, \cdots, Se_n)$ 是规范正交的, 这就证明了 (c) 蕴含 (d).

显然 (d) 蕴含 (e).

现在假设 (e) 成立. 设 $(e_1, \cdots, e_n)$ 是 V 的规范正交基使得 $(Se_1, \cdots, Se_n)$ 是规范正交的. 若 $\boldsymbol{v} \in V$, 则

$$\begin{aligned}\|Sv\|^2 &= \|S(\langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n)\|^2 \\ &= \|\langle v, e_1\rangle Se_1 + \cdots + \langle v, e_n\rangle Se_n\|^2 \\ &= |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 \\ &= \|v\|^2,\end{aligned}$$

其中第一个和最后一个等式由 6.17 推出. 开方即知 S 是等距同构, 这就证明了 (e) 蕴含 (a).

已证 (a)⇒ (b)⇒ (c)⇒ (d)⇒ (e)⇒ (a), 于是, 从 (a) 到 (e) 都互相等价. 用 S^* 代替 S 可知, 从 (f) 到 (j) 都互相等价. 因此, 为了完成证明, 只需证明从 (a) 到 (e) 的条件之一等价于从 (f) 到 (j) 的条件之一. 联系这两组条件最容易的办法就是证明 (c) 等价于 (h). 一般地, S 当然不必和 S^* 交换. 然而, $S^*S = I$ 当且仅当 $SS^* = I$; 这是第 3 章习题 23 的一种特殊情况. 于是 (c) 等价于 (h). ■

上面这个定理证明了每个等距同构都是正规的 (参见 7.36 中的 (a), (c), (h)). 于是, 正规算子的刻画可以用来给出等距同构的完全描述. 我们要在以下两个定理中做这件事.

7.37 定理: 设 V 是复内积空间, $S \in \mathcal{L}(V)$, 则 S 是等距同构当且仅当 S 的本征值的绝对值都是 1, 而且 V 有一个由 S 的本征向量组成的规范正交基.

证明: 我们已经证明 (参见本节第一段), 如果 S 的本征值的绝对值都是 1, 而且 V 有一个由 S 的本征向量组成的规范正交基, 则 S 是等距同构.

为了证明另一个方面, 假设 S 是等距同构. 由复谱定理 (7.9), V 有一个由 S 的本征向量组成的规范正交基 $(e_1, \cdots, e_n)$. 对于 $j \in \{1, \cdots, n\}$, 设 λ_j 是相应于 e_j 的本征值, 则

$$|\lambda_j| = \|\lambda_j e_j\| = \|Se_j\| = \|e_j\| = 1.$$

于是, S 的每个本征值的绝对值都是 1. ■

如果 $\theta \in \mathbf{R}$, 那么你应该验证, 在 $\mathbf{R}^2$ 上 (以原点为中心) 逆时针旋转 θ 角度所确定的算子关于标准基有矩阵 7.39. 下一个结果表明实内积空间上的等距同构由以下三部分组成: 一部分类似于 2 维子空间上的旋转, 一部分等于恒等算子, 一部分相当于乘以 -1.

7.38 定理: *假设 V 是实内积空间, $S \in \mathcal{L}(V)$, 则 S 是等距同构当且仅当 V 有一个规范正交基使得关于此基 S 有分块对角矩阵, 其中对角线上的每个块都是由 1 或 -1 构成的 1×1 矩阵, 或者是如下形式的 2×2 矩阵*

这个定理意味着奇数维实内积空间上的等距同构一定有本征值 1 或 -1.

7.39
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \theta \in (0, \pi).$$

证明: 首先假设 S 是等距同构. 因为 S 是正规的, 所以由 7.25 可知, V 有规范正交基使得 S 关于此基有分块对角矩阵, 其中每个块都是 1×1 矩阵或如下形式的 2×2 矩阵

7.40
$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad b > 0.$$

如果 λ 是 S (关于上述基) 的矩阵对角线上 1×1 矩阵的元素, 那么存在基向量 $\boldsymbol{e}_j$ 使得 $S\boldsymbol{e}_j = \lambda \boldsymbol{e}_j$. 因为 S 是等距同构, 所以 $|\lambda| = 1$. 于是, $\lambda = 1$ 或 $\lambda = -1$, 因为只有这两个实数的绝对值是 1.

现在考虑位于 S 的矩阵的对角线上形如 7.40 的 2×2 矩阵. 有基向量 $\boldsymbol{e}_j, \boldsymbol{e}_{j+1}$ 使得

$$S\boldsymbol{e}_j = a\boldsymbol{e}_j + b\boldsymbol{e}_{j+1}.$$

于是

$$1 = \|\boldsymbol{e}_j\|^2 = \|S\boldsymbol{e}_j\|^2 = a^2 + b^2.$$

再由 $b > 0$ 知, 存在 $\theta \in (0, \pi)$ 使得 $a = \cos\theta$, $b = \sin\theta$.

因而矩阵 7.40 具有所要求的形式 7.39, 这就完成了这个方向的证明.

反之, 假设 V 有规范正交基使得 S 关于此基的矩阵具有定理所要求的形式. 于是有直和分解

$$V = U_1 \oplus \cdots \oplus U_m,$$

其中每个 U_j 都是 V 的 1 维或 2 维子空间. 进一步, 任意两个属于不同 U_j 的向量都是正交的, 并且每个 $S|_{U_j}$ 都是把 U_j 映到 U_j 内的等距同构. 若 $\boldsymbol{v} \in V$, 则有

$$\boldsymbol{v} = \boldsymbol{u}_1 + \cdots + \boldsymbol{u}_m,$$

其中 $\boldsymbol{u}_j \in U_j$. 把 S 应用到上式并取范数可得

$$\begin{aligned}\|S\boldsymbol{v}\|^2 &= \|S\boldsymbol{u}_1 + \cdots + S\boldsymbol{u}_m\|^2 \\ &= \|S\boldsymbol{u}_1\|^2 + \cdots + \|S\boldsymbol{u}_m\|^2 \\ &= \|\boldsymbol{u}_1\|^2 + \cdots + \|\boldsymbol{u}_m\|^2 \\ &= \|\boldsymbol{v}\|^2.\end{aligned}$$

于是, S 是等距同构. ■

§7.6　极分解与奇异值分解

回想一下我们在 $\mathbf{C}$ 和 $\mathcal{L}(V)$ 之间做的类比. 按照这个类比, 一个复数 z 相应于一个算子 T, 而 $\bar{z}$ 相应于 T^*. 实数相应于自伴算子, 而非负数相应于 (不恰当地所谓) 正算子. $\mathbf{C}$ 的另一个重要的子集是单位圆, 它由所有满足 $|z| = 1$ 的复数 z 组成. 条件 $|z| = 1$ 等价于 $\bar{z}z = 1$. 按照我们的类比, 这相应于条件 $T^*T = I$, 等价于 T 是等距同构 (参见 7.36). 也就是说, $\mathbf{C}$ 中的单位圆相应于全体等距同构.

继续我们的类比, 注意到每个非零复数 z 都可以写成如下形式

$$z = \left(\frac{z}{|z|}\right)|z| = \left(\frac{z}{|z|}\right)\sqrt{\bar{z}z},$$

其中第一个因子, 即 $z/|z|$, 是单位圆中的元素. 这种类比使我们

猜到, 任何算子 $T \in \mathcal{L}(V)$ 都可以写成一个等距同构和 $\sqrt{T^*T}$ 的乘积. 我们现在就来证明这个猜测确实是对的.

7.41 极分解 (Polar Decomposition): 如果 $T \in \mathcal{L}(V)$, 则有一个等距同构 $S \in \mathcal{L}(V)$ 使得

$$T = S\sqrt{T^*T}.$$

如果你知道一点复分析, 那么你就会认识到与复数的极坐标的类比: 每个复数都可以写成 $\mathrm{e}^{\theta i}r$, 其中 $\theta \in [0, 2\pi)$, 而且 $r \geqslant 0$. 注意到 $\mathrm{e}^{\theta i}$ 在单位圆上相应于 S 是等距同构, 而 r 是非负的相应于 T^*T 是正算子.

证明: 假设 $T \in \mathcal{L}(V)$. 若 $\boldsymbol{v} \in V$, 则

$$\begin{aligned}\|T\boldsymbol{v}\|^2 &= \langle T\boldsymbol{v}, T\boldsymbol{v}\rangle \\ &= \langle T^*T\boldsymbol{v}, \boldsymbol{v}\rangle \\ &= \langle \sqrt{T^*T}\sqrt{T^*T}\boldsymbol{v}, \boldsymbol{v}\rangle \\ &= \langle \sqrt{T^*T}\boldsymbol{v}, \sqrt{T^*T}\boldsymbol{v}\rangle \\ &= \|\sqrt{T^*T}\boldsymbol{v}\|^2.\end{aligned}$$

于是, 对所有 $\boldsymbol{v} \in V$ 都有

7.42
$$\|T\boldsymbol{v}\| = \|\sqrt{T^*T}\boldsymbol{v}\|.$$

定义线性映射 $S_1 : \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$ 如下

7.43
$$S_1(\sqrt{T^*T}\boldsymbol{v}) = T\boldsymbol{v}.$$

证明的要点是把 S_1 扩张成一个等距同构 $S \in \mathcal{L}(V)$, 使得 $T = S\sqrt{T^*T}$. 现在来看具体证明.

首先必须验证 S_1 的定义是合理的. 为此, 假设 $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ 使得 $\sqrt{T^*T}\boldsymbol{v}_1 = \sqrt{T^*T}\boldsymbol{v}_2$. 为使 7.43 给出的定义有意义, 我们必须证明 $T\boldsymbol{v}_1 = T\boldsymbol{v}_2$. 然而,

$$\begin{aligned}\|T\boldsymbol{v}_1 - T\boldsymbol{v}_2\| &= \|T(\boldsymbol{v}_1 - \boldsymbol{v}_2)\| \\ &= \|\sqrt{T^*T}(\boldsymbol{v}_1 - \boldsymbol{v}_2)\| \\ &= \|\sqrt{T^*T}\boldsymbol{v}_1 - \sqrt{T^*T}\boldsymbol{v}_2\| \\ &= 0,\end{aligned}$$

其中第二个等式成立是由于 7.42. 上式证明了 $T\boldsymbol{v}_1 = T\boldsymbol{v}_2$, 故

在余下的证明中，我们要做的就是把 S_1 扩张成整个 V 上的一个等距同构 S.

S_1 的确是定义合理的. 你应该验证 S_1 是线性映射.

由 7.43 可知, S_1 把 $\operatorname{range}\sqrt{T^*T}$ 映到 $\operatorname{range}T$ 上. 显然, 由 7.42 和 7.43 可知, 对所有 $\boldsymbol{u}\in\operatorname{range}\sqrt{T^*T}$ 都有 $\|S_1\boldsymbol{u}\|=\|\boldsymbol{u}\|$. 特别地, S_1 是单的. 于是, 对 S_1 应用 3.4 可得

$$\dim\operatorname{range}\sqrt{T^*T}=\dim\operatorname{range}T.$$

这意味着 $\dim(\operatorname{range}\sqrt{T^*T})^\perp=\dim(\operatorname{range}T)^\perp$ (参见第 6 章习题 15). 于是, 可取 $(\operatorname{range}\sqrt{T^*T})^\perp$ 的规范正交基 $(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_m)$ 及 $(\operatorname{range}T)^\perp$ 的规范正交基 $(\boldsymbol{f}_1,\cdots,\boldsymbol{f}_m)$; 关键是这两个规范正交基的长度相同. 定义线性映射 $S_2:(\operatorname{range}\sqrt{T^*T})^\perp\to(\operatorname{range}T)^\perp$ 如下

$$S_2(a_1\boldsymbol{e}_1+\cdots+a_m\boldsymbol{e}_m)=a_1\boldsymbol{f}_1+\cdots+a_m\boldsymbol{f}_m.$$

显然, 对所有 $\boldsymbol{w}\in(\operatorname{range}\sqrt{T^*T})^\perp$ 都有 $\|S_2\boldsymbol{w}\|=\|\boldsymbol{w}\|$.

现在设 S 是 V 上的算子, 并设它在 $\operatorname{range}\sqrt{T^*T}$ 上和 S_1 相等, 而在 $(\operatorname{range}\sqrt{T^*T})^\perp$ 上和 S_2 相等. 更准确地说, 因为每个 $\boldsymbol{v}\in V$ 都可以唯一地写成

7.44
$$\boldsymbol{v}=\boldsymbol{u}+\boldsymbol{w},$$

其中 $\boldsymbol{u}\in\operatorname{range}\sqrt{T^*T},\boldsymbol{w}\in(\operatorname{range}\sqrt{T^*T})^\perp$ (参见 6.29), 所以按照上述分解, 对于 $\boldsymbol{v}\in V$, 可以定义 $S\boldsymbol{v}$ 如下

$$S\boldsymbol{v}=S_1\boldsymbol{u}+S_2\boldsymbol{w}.$$

对每个 $\boldsymbol{v}\in V$ 都有

$$S(\sqrt{T^*T}\boldsymbol{v})=S_1(\sqrt{T^*T}\boldsymbol{v})=T\boldsymbol{v}$$

故 $T=S\sqrt{T^*T}$. 还需要证明 S 是等距同构. 然而, 这很容易由勾股定理得到: 若 $\boldsymbol{v}\in V$ 已经分解成 7.44 的形式, 则

$$\begin{aligned}\|S\boldsymbol{v}\|^2&=\|S_1\boldsymbol{u}+S_2\boldsymbol{w}\|^2\\&=\|S_1\boldsymbol{u}\|^2+\|S_2\boldsymbol{w}\|^2\\&=\|\boldsymbol{u}\|^2+\|\boldsymbol{w}\|^2\\&=\|\boldsymbol{v}\|^2,\end{aligned}$$

其中第二个等式成立是由于 $S_1\boldsymbol{u}\in\operatorname{range}T$, $S_2\boldsymbol{w}\in(\operatorname{range}T)^\perp$.

■

极分解 (7.41) 说的是, V 上的每个算子都是一个等距同构和一个正算子的乘积. 于是, V 上的每个算子都可以写成两个算子之积, 这两个算子都来自我们已经完全描述, 并且能够比较好地理解的算子类. 7.37 和 7.38 描述了等距同构; 谱定理 (7.9 和 7.13) 描述了正算子 (它们都是自伴的).

具体地, 设 $T = S\sqrt{T^*T}$ 是 $T \in \mathcal{L}(V)$ 的极分解, 其中 S 是等距同构, 则 V 有规范正交基使得 S 关于此基有对角矩阵 (若 $\mathbf{F} = \mathbf{C}$), 或者有分块对角矩阵, 其中每个块的大小都不超过 2×2 (若 $\mathbf{F} = \mathbf{R}$), 并且 V 还有规范正交基使得 $\sqrt{T^*T}$ 关于此基有对角矩阵. 要注意: 未必有规范正交基使得 S 和 $\sqrt{T^*T}$ 的矩阵同时具有这么好的形式 (对角矩阵或由较小的块组成的分块对角矩阵). 也就是说, S 需要一个规范正交基, 而 $\sqrt{T^*T}$ 可能需要另一个不同的规范正交基.

设 $T \in \mathcal{L}(V)$. T 的**奇异值 (singular value)** 就是 $\sqrt{T^*T}$ 的本征值, 而且每个本征值 λ 都要重复 $\dim \operatorname{null}(T^*T - \lambda I)$ 次. 因为 T 的奇异值都是正算子 $\sqrt{T^*T}$ 的本征值, 所以它们都非负.

例如, 若 $T \in \mathcal{L}(\mathbf{F}^4)$ 定义如下

7.45 $$T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4),$$

那么你应该验证, $T^*T(z_1, z_2, z_3, z_4) = (9z_1, 4z_2, 0, 9z_4)$. 于是

$$\sqrt{T^*T}(z_1, z_2, z_3, z_4) = (3z_1, 2z_2, 0, 3z_4),$$

并且 $\sqrt{T^*T}$ 的本征值是 $3, 2, 0$. 显然,

$$\dim \operatorname{null}(\sqrt{T^*T} - 3I) = 2, \quad \dim \operatorname{null}(\sqrt{T^*T} - 2I) = 1,$$

$$\dim \operatorname{null} \sqrt{T^*T} = 1.$$

因此, T 的奇异值是 $3, 3, 2, 0$. 你应该验证, 在这个例子中, T 的本征值只有 -3 和 0.

把谱定理和 5.21 (尤其是 (e)) 用于正算子 (故自伴) $\sqrt{T^*T}$ 可知, 每个 $T \in \mathcal{L}(V)$ 都有 $\dim V$ 个奇异值. 例如, 我们在前一段看到, 在 4 维向量空间 $\mathbf{F}^4$ 上, 由 7.45 定义的算子 T 有 4 个奇异值 (它们是 $3, 3, 2, 0$).

下一个结果表明, V 上的每个算子都可以通过它的奇异值和 V 的两个规范正交基很好地描述.

7.46 奇异值分解 (Singular-Value Decomposition)： 设 $T \in \mathcal{L}(V)$ 有奇异值 $s_1, \cdots, s_n$, 则有 V 的两个规范正交基 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ 和 $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ 使得对每个 $\boldsymbol{v} \in V$ 都有

7.47 $$T\boldsymbol{v} = s_1\langle \boldsymbol{v}, \boldsymbol{e}_1\rangle \boldsymbol{f}_1 + \cdots + s_n\langle \boldsymbol{v}, \boldsymbol{e}_n\rangle \boldsymbol{f}_n.$$

证明： 对 $\sqrt{T^*T}$ 应用谱定理 (也参看 7.14), 则 V 有规范正交基 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ 使得 $\sqrt{T^*T}\boldsymbol{e}_j = s_j\boldsymbol{e}_j$, $j = 1, \cdots, n$. 对每个 $\boldsymbol{v} \in V$ 都有 (参见 6.17)

$$\boldsymbol{v} = \langle \boldsymbol{v}, \boldsymbol{e}_1\rangle \boldsymbol{e}_1 + \cdots + \langle \boldsymbol{v}, \boldsymbol{e}_n\rangle \boldsymbol{e}_n.$$

把 $\sqrt{T^*T}$ 作用到这个等式的两端, 则对每个 $\boldsymbol{v} \in V$ 都有

$$\sqrt{T^*T}\boldsymbol{v} = s_1\langle \boldsymbol{v}, \boldsymbol{e}_1\rangle \boldsymbol{e}_1 + \cdots + s_n\langle \boldsymbol{v}, \boldsymbol{e}_n\rangle \boldsymbol{e}_n.$$

这个证明就是极分解用途的例证.

由极分解 (7.41), 有等距同构 $S \in \mathcal{L}(V)$ 使得 $T = S\sqrt{T^*T}$. 把 S 作用到上式两端, 则对每个 $\boldsymbol{v} \in V$ 都有

$$T\boldsymbol{v} = s_1\langle \boldsymbol{v}, \boldsymbol{e}_1\rangle S\boldsymbol{e}_1 + \cdots + s_n\langle \boldsymbol{v}, \boldsymbol{e}_n\rangle S\boldsymbol{e}_n.$$

对每个 j, 令 $\boldsymbol{f}_j = S\boldsymbol{e}_j$. 因为 S 是等距同构, 所以 $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ 是 V 的规范正交基 (参见 7.36). 上式现在就变成

$$T\boldsymbol{v} = s_1\langle \boldsymbol{v}, \boldsymbol{e}_1\rangle \boldsymbol{f}_1 + \cdots + s_n\langle \boldsymbol{v}, \boldsymbol{e}_n\rangle \boldsymbol{f}_n, \quad \boldsymbol{v} \in V.$$ ■

我们在研究从一个向量空间到另一个向量空间的线性映射时, 讨论了线性映射关于第一个向量空间的基和第二个向量空间的基的矩阵. 在研究算子 (即从一个向量空间到其自身的线性映射) 时, 我们几乎总是让同一个基扮演这两种角色.

奇异值分解给了我们一个少有的机会：讨论算子关于两个不同基的矩阵. 为此, 设 $T \in \mathcal{L}(V)$, $s_1, \cdots, s_n$ 表示 T 的所有奇异值, 并设 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ 和 $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ 都是 V 的规范正交基使得奇异值分解 7.47 成立, 则显然有

$$\mathcal{M}(T,(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n),(\boldsymbol{f}_1,\cdots,\boldsymbol{f}_n))=\begin{bmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{bmatrix}.$$

也就是说, 只要允许我们在处理算子时使用两个不同的基, 而不是像通常那样只使用单独的一个基, 那么 V 上每个算子关于 V 的某些规范正交基就都有对角矩阵.

奇异值和奇异值分解有很多应用 (习题中给出了一些), 包括在计算线性代数中的应用. 为了计算一个算子的奇异值的数值近似值, 首先计算 T^*T, 然后计算 T^*T 的近似本征值 (计算正算子的近似本征值有很好的技术). T^*T 的这些 (近似) 本征值的非负平方根就是 T 的 (近似) 奇异值 (正如在 7.28 的证明中所见). 也就是说, 不需要计算 T^*T 的平方根, 就能得出 T 的近似奇异值.

习 题

1. 按照下面的定义, $\mathcal{P}_2(\mathbf{R})$ 是内积空间,

$$\langle p, q\rangle = \int_0^1 p(x)q(x)\mathrm{d}x.$$

定义 $T \in \mathcal{L}(\mathcal{P}_2(\mathbf{R}))$ 使得 $T(a_0 + a_1x + a_2x^2) = a_1x$.

(a) 证明 T 不是自伴的.

(b) T 关于基 $(1, x, x^2)$ 的矩阵是

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

虽然 T 不是自伴的, 但是这个矩阵却和它的共轭转置相等. 解释为什么这并不矛盾.

2. 证明或举反例: 有限维内积空间上的两个自伴算子之积是自伴的.

3. (a) 证明: 若 V 是实内积空间, 则 V 上的自伴算子之集是 $\mathcal{L}(V)$ 的子空间.

(b) 证明: 若 V 是复内积空间, 则 V 上的自伴算子之集不是 $\mathcal{L}(V)$ 的子空间.

4. 设 $P \in \mathcal{L}(V)$ 使得 $P^2 = P$. 证明 P 是正交投影当且仅当 P 是自伴的.

5. 证明: 若 $\dim V \geqslant 2$, 则 V 上的正规算子之集不是 $\mathcal{L}(V)$ 的子空间.

6. 证明: 若 $T \in \mathcal{L}(V)$ 是正规的, 则

$$\operatorname{range} T = \operatorname{range} T^*.$$

7. 证明: 若 $T \in \mathcal{L}(V)$ 是正规的, 则对每个正整数 k 都有

$$\operatorname{null} T^k = \operatorname{null} T, \quad \operatorname{range} T^k = \operatorname{range} T.$$

8. 证明没有自伴算子 $T \in \mathcal{L}(\mathbf{R}^3)$ 能使得

$$T(1,2,3) = (0,0,0), \quad T(2,5,7) = (2,5,7).$$

9. 证明：在复内积空间上，一个正规算子是自伴的当且仅当它的所有本征值都是实的.

习题 9 (对于正规算子) 强化了自伴算子与实数之间的类比.

10. 设 V 是复内积空间，$T \in \mathcal{L}(V)$ 是正规算子使得 $T^9 = T^8$. 证明 T 是自伴的，并且 $T^2 = T$.

11. 设 V 是复内积空间. 证明 V 上每个正规算子都有平方根. (算子 $S \in \mathcal{L}(V)$ 称为 $T \in \mathcal{L}(V)$ 的**平方根 (square root)**，如果 $S^2 = T$.)

12. 给出实内积空间 V，算子 $T \in \mathcal{L}(V)$ 以及满足 $\alpha^2 < 4\beta$ 的实数 α, β，使得 $T^2 + \alpha T + \beta I$ 不可逆.

这个习题表明，即使对于实向量空间，7.11 中 T 自伴的假设也是必要的.

13. 证明或举反例：V 上每个自伴算子都有立方根. (算子 $S \in \mathcal{L}(V)$ 称为 $T \in \mathcal{L}(V)$ 的**立方根 (cube root)**，如果 $S^3 = T$.)

14. 设 $T \in \mathcal{L}(V)$ 是自伴的，$\lambda \in \mathbf{F}$，$\epsilon > 0$. 证明，若有 $v \in V$ 使得 $\|v\| = 1$，并且

$$\|T\boldsymbol{v} - \lambda \boldsymbol{v}\| < \epsilon,$$

则 T 有本征值 λ' 使得 $|\lambda - \lambda'| < \epsilon$.

15. 设 U 是有限维实向量空间，并且 $T \in \mathcal{L}(U)$. 证明 U 有一个由 T 的本征向量组成的基当且仅当存在 U 上的内积使得 T 是自伴算子.

16. 给出一个内积空间上的算子 T 使得 T 有一个不变子空间，但是它的正交补在 T 下却不是不变的.

这个习题说明，如果没有 T 正规的假设，7.18 就不成立.

17. 证明 V 上的两个正算子之和是正的.

18. 证明：若 $T \in \mathcal{L}(V)$ 是正的，则对每个正整数 k，T^k 都是正的.

19. 设 T 是 V 上的正算子. 证明 T 可逆当且仅当对每个 $\boldsymbol{v} \in V \backslash \{\mathbf{0}\}$ 都有

$$\langle T\boldsymbol{v}, \boldsymbol{v}\rangle > 0.$$

20. 证明或反驳: $\mathbf{F}^2$ 上的恒等算子有无穷多个自伴的平方根.

21. 证明或举反例: 若 $S \in \mathcal{L}(V)$, 并且 V 有规范正交基 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ 使得对每个 $\boldsymbol{e}_j$ 都有 $\|S\boldsymbol{e}_j\| = 1$, 则 S 是等距同构.

22. 证明: 若 $S \in \mathcal{L}(\mathbf{R}^3)$ 是等距同构, 则有非零向量 $\boldsymbol{x} \in \mathbf{R}^3$ 使得 $S^2\boldsymbol{x} = \boldsymbol{x}$.

23. 定义 $T \in \mathcal{L}(\mathbf{F}^3)$ 如下

$$T(z_1, z_2, z_3) = (z_3, 2z_1, 3z_2).$$

求出一个等距同构 $S \in \mathcal{L}(\mathbf{F}^3)$ 使得 $T = S\sqrt{T^*T}$.

习题 24 表明, 如果把 T 写成等距同构与一个正算子之积 (犹如在极分解中), 则此正算子必为 $\sqrt{T^*T}$.

24. 设 $T \in \mathcal{L}(V)$, $S \in \mathcal{L}(V)$ 都是等距同构, 并且 $R \in \mathcal{L}(V)$ 是正算子使得 $T = SR$. 证明 $R = \sqrt{T^*T}$.

25. 设 $T \in \mathcal{L}(V)$. 证明 T 可逆当且仅当有唯一一个等距同构 $S \in \mathcal{L}(V)$ 使得 $T = S\sqrt{T^*T}$.

26. 证明: 若 $T \in \mathcal{L}(V)$ 是自伴的, 则 T 的奇异值等于 T 的本征值的绝对值 (适当重复).

27. 证明或举反例: 若 $T \in \mathcal{L}(V)$, 则 T^2 的奇异值等于 T 的奇异值的平方.

28. 设 $T \in \mathcal{L}(V)$. 证明 T 可逆当且仅当 0 不是 T 的奇异值.

29. 设 $T \in \mathcal{L}(V)$. 证明 $\dim \operatorname{range} T$ 等于 T 的非零奇异值的个数.

30. 设 $S \in \mathcal{L}(V)$. 证明 S 是等距同构当且仅当 S 的所有奇异值都等于 1.

31. 设 $T_1, T_2 \in \mathcal{L}(V)$. 证明 T_1 和 T_2 有同样的奇异值当且仅当有等距同构 $S_1, S_2 \in \mathcal{L}(V)$ 使得 $T_1 = S_1T_2S_2$.

32. 设 $T \in \mathcal{L}(V)$ 有奇异值分解

$$T\boldsymbol{v} = s_1\langle \boldsymbol{v}, \boldsymbol{e}_1\rangle \boldsymbol{f}_1 + \cdots + s_n\langle \boldsymbol{v}, \boldsymbol{e}_n\rangle \boldsymbol{f}_n, \quad \boldsymbol{v} \in V,$$

其中 $s_1, \cdots, s_n$ 是 T 的所有奇异值, $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ 和 $(\boldsymbol{f}_1, \cdots, \boldsymbol{f}_n)$ 都是 V 的规范正交基.

(a) 证明：对每个 $\boldsymbol{v} \in V$ 都有

$$T^*\boldsymbol{v} = s_1\langle \boldsymbol{v}, \boldsymbol{f}_1\rangle \boldsymbol{e}_1 + \cdots + s_n\langle \boldsymbol{v}, \boldsymbol{f}_n\rangle \boldsymbol{e}_n.$$

(b) 证明：若 T 可逆, 则对每个 $\boldsymbol{v} \in V$ 都有

$$T^{-1}\boldsymbol{v} = \frac{\langle \boldsymbol{v}, \boldsymbol{f}_1\rangle \boldsymbol{e}_1}{s_1} + \cdots + \frac{\langle \boldsymbol{v}, \boldsymbol{f}_n\rangle \boldsymbol{e}_n}{s_n}.$$

33. 设 $T \in \mathcal{L}(V)$. 令 $\hat{s}$ 表示 T 的最小奇异值, s 表示 T 的最大奇异值. 证明对每个 $\boldsymbol{v} \in V$ 都有

$$\hat{s}\|\boldsymbol{v}\| \leqslant \|T\boldsymbol{v}\| \leqslant s\|\boldsymbol{v}\|.$$

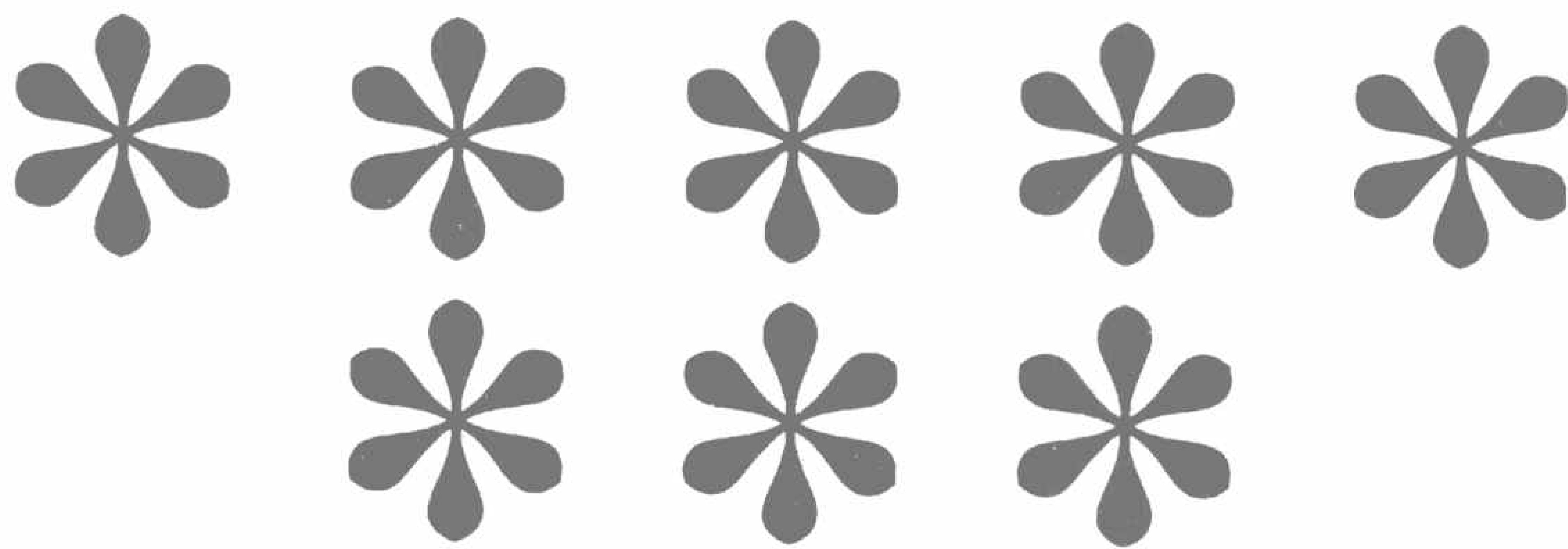
34. 设 $T', T'' \in \mathcal{L}(V)$. 令 s' 表示 T' 的最大奇异值, s'' 表示 T'' 的最大奇异值, s 表示 $T' + T''$ 的最大奇异值. 证明 $s \leqslant s' + s''$.

第 8 章　复向量空间上的算子

本章将更深入地研究复向量空间上算子的结构，这里用不到内积，因此我们又回到了一般的有限维向量空间 (而不是特殊的内积空间) 的情形. 于是本章作如下假设:

$\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$.
V 是 $\mathbf{F}$ 上的有限维非零向量空间.

本章中的有些结果在实向量空间上也成立，因而我们并未仅仅假设 V 是复向量空间. 本章的一些结果只对复向量空间的情形给出了证明，我们将在下一章给出其中大部分结果在实向量空间上的类似结果的证明. 因为复向量空间上的证明通常比实向量空间上的类似证明要简单，所以我们首先处理复向量空间.



§8.1　广义本征向量

有些算子因为没有足够多的本征向量而没有一个好的描述. 因此, 本节将引入广义本征向量的概念, 这一概念对算子结构的描述起着重要作用.

为了理解为什么本征向量不够多, 我们来考察如何通过算子定义域中的不变子空间分解来描述这个算子. 取定 $T \in \mathcal{L}(V)$, 为了描述 T, 我们想要找到一个 "好的" 直和分解

8.1
$$V = U_1 \oplus \cdots \oplus U_m,$$

其中每个 U_j 都是 V 在 T 下的不变子空间. 可能存在的、最简单的非零不变子空间是一维的. 分解 8.1 中的每个 U_j 都是在 T 下不变的一维子空间, 当且仅当 V 有一个由 T 的本征向量组成的基 (参见 5.21), 当且仅当 V 有如下分解

8.2
$$V = \operatorname{null}(T - \lambda_1 I) \oplus \cdots \oplus \operatorname{null}(T - \lambda_m I),$$

其中 $\lambda_1, \cdots, \lambda_m$ 是 T 的所有互不相同的本征值 (参见 5.21).

上一章我们证明了, 内积空间上的每个自伴算子都有形如 8.2 的分解 (参见 7.14). 可惜的是, 即使在复向量空间上, 形如 8.2 的分解对更一般的算子也可能不成立. 5.19 所给出的算子就是一个例子, 它没有足够多的本征向量使得 8.2 成立. 我们现在要引入的广义本征向量将会弥补这种不足. 本章的主要目标是要证明: 如果 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$, 那么

$$V = \operatorname{null}(T - \lambda_1 I)^{\dim V} \oplus \cdots \oplus \operatorname{null}(T - \lambda_m I)^{\dim V},$$

其中 $\lambda_1, \cdots, \lambda_m$ 是 T 的所有互不相同的本征值 (参见 8.23).

设 $T \in \mathcal{L}(V)$, 并且 λ 是 T 的本征值. 对于向量 $v \in V$, 如果存在正整数 j 使得

8.3
$$(T - \lambda I)^j v = 0,$$

则称 v 是 T 的相应于 λ 的**广义本征向量 (generalized eigenvector)**. 注意, T 的每个本征向量都是 T 的广义本征向量 (在上面的等式中取 $j = 1$), 但反之不成立. 例如, 如果 $T \in \mathcal{L}(\mathbf{C}^3)$ 定义为

$$T(z_1, z_2, z_3) = (z_2, 0, z_3),$$

那么, 对所有 $z_1, z_2 \in \mathbf{C}$ 都有 $T^2(z_1, z_2, 0) = 0$. 因此 $\mathbf{C}^3$ 中最后一个坐标等于 0 的元素都是 T 的广义本征向量. 你应该验证

$$\mathbf{C}^3 = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{C}\} \oplus \{(0, 0, z_3) : z_3 \in \mathbf{C}\},$$

其中右端的第一个子空间等于该算子相应于本征值 0 的广义本征向量之集, 第二个子空间等于该算子相应于本征值 1 的广义本征向量之集. 我们将在本章后面证明, 复向量空间上的每个算子都有一个由广义本征向量所给出的分解 (参见 8.23).

虽然广义本征向量定义中的 j 可以是任意正数, 但是我们很快就会看到, 每个广义本征向量都满足形如 8.3 的方程而且其中的 j 等于 V 的维数. 要证明这一点, 现在我们来讨论算子幂的零空间.

注意, 我们没有定义广义本征值的概念, 因为这样不会得到任何新东西. 理由是: 如果对某个正整数 j, $(T - \lambda I)^j$ 不是单的, 那么 $T - \lambda I$ 也不是单的, 因此 λ 是 T 的本征值.

设 $T \in \mathcal{L}(V)$, 并且 k 是非负整数. 如果 $T^k \boldsymbol{v} = \mathbf{0}$, 那么 $T^{k+1}\boldsymbol{v} = T(T^k\boldsymbol{v}) = T(\mathbf{0}) = \mathbf{0}$. 因此 $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$. 也就是说,

8.4 $\{0\} = \operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^k \subset \operatorname{null} T^{k+1} \subset \cdots.$

下一个命题是说, 如果在这个子空间序列中有相邻的两项相等, 那么此后的所有项都相等.

8.5 命题: 如果 $T \in \mathcal{L}(V)$, 并且 m 是一个非负整数使得 $\operatorname{null} T^m = \operatorname{null} T^{m+1}$, 那么

$$\operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \cdots.$$

证明: 设 $T \in \mathcal{L}(V)$, m 是一个非负整数使得 $\operatorname{null} T^m = \operatorname{null} T^{m+1}$, 并设 k 是正整数. 往证

$$\operatorname{null} T^{m+k} = \operatorname{null} T^{m+k+1}.$$

我们已经知道 $\operatorname{null} T^{m+k} \subset \operatorname{null} T^{m+k+1}$. 要证明另一个方向的包含关系, 设 $\boldsymbol{v} \in \operatorname{null} T^{m+k+1}$, 则

$$0 = T^{m+k+1}\boldsymbol{v} = T^{m+1}(T^k\boldsymbol{v}).$$

因此

$$T^k\boldsymbol{v} \in \text{null } T^{m+1} = \text{null } T^m.$$

于是 $\mathbf{0} = T^m(T^k\boldsymbol{v}) = T^{m+k}\boldsymbol{v}$, 即 $\boldsymbol{v} \in \text{null } T^{m+k}$. 由此得 $\text{null } T^{m+k+1} \subset \text{null } T^{m+k}$. ■

上面的命题引出一个问题: 是否存在非负整数 m, 使得 $\text{null } T^m = \text{null } T^{m+1}$? 下面的命题表明, 这个等式至少在 m 等于 T 的定义域的维数时成立.

8.6 命题: 如果 $T \in \mathcal{L}(V)$, 那么

$$\text{null } T^{\dim V} = \text{null } T^{\dim V+1} = \text{null } T^{\dim V+2} = \cdots.$$

证明: 设 $T \in \mathcal{L}(V)$. 只需证明 $\text{null } T^{\dim V} = \text{null } T^{\dim V+1}$ (由 8.5). 假设不然, 则由 8.5 得

$$\{0\} = \text{null } T^0 \subsetneq \text{null } T^1 \subsetneq \cdots \subsetneq \text{null } T^{\dim V} \subsetneq \text{null } T^{\dim V+1},$$

其中符号 $\subsetneq$的意思是 "包含于, 但是不等于". 在上述链中的每个严格包含关系处, 维数至少增加 1. 因此 $\dim \text{null } T^{\dim V+1} \geqslant \dim V+1$, 矛盾, 这是因为 V 的子空间的维数不可能大于 $\dim V$. ■

现在我们得到了许诺过的关于广义本征向量的描述.

这个推论说明, $T \in \mathcal{L}(V)$ 相应于本征值 λ 的广义本征向量之集是 V 的子空间.

8.7 推论: 设 $T \in \mathcal{L}(V)$, 并且 λ 是 T 的本征值, 则 T 的相应于 λ 的广义本征向量之集等于 $\text{null } (T-\lambda I)^{\dim V}$.

证明: 若 $\boldsymbol{v} \in \text{null } (T-\lambda I)^{\dim V}$, 则 $\boldsymbol{v}$ 显然是 T 的相应于 λ 的广义本征向量 (由广义本征向量的定义).

反之, 设 $\boldsymbol{v} \in V$ 是 T 相应于 λ 的广义本征向量, 则存在正整数 j 使得

$$\boldsymbol{v} \in \text{null } (T-\lambda I)^j.$$

由 8.5 和 8.6 (以 $T-\lambda I$ 代替 T) 得 $\boldsymbol{v} \in \text{null } (T-\lambda I)^{\dim V}$. ■

一个算子称为**幂零的 (nilpotent)**, 如果它的某个幂等于 0. 例如, 下面定义的算子 $N \in \mathcal{L}(\mathbf{F}^4)$

$$N(z_1, z_2, z_3, z_4) = (z_3, z_4, 0, 0)$$

是幂零的, 因为 $N^2 = 0$. 又如, $\mathcal{P}_m(\mathbf{R})$ 上的微分算子是幂零的, 这是因为任何次数不超过 m 的多项式的 $m+1$ 阶导数都等于 0. 注意, 在这个 $m+1$ 维空间上, 我们需要把幂零算子自乘到 $m+1$ 次幂才能得到 0. 下一个推论表明, 不需要用到比空间的维数更高的幂.

拉丁词 nil 的意思是无或者零; 拉丁词 potent 的意思是幂. 于是 nilpotent 的字面意思是幂为零.

8.8 推论: 设 $N \in \mathcal{L}(V)$ 是幂零的, 则 $N^{\dim V} = 0$.

证明: 因为 N 是幂零的, 所以 V 中每个向量都是相应于本征值 0 的广义本征向量. 由此利用 8.7 即得 $\operatorname{null} N^{\dim V} = V$. ■

前面讨论了算子幂的零空间, 现在考虑值域. 设 $T \in \mathcal{L}(V)$, 并且 k 是非负整数. 如果 $\boldsymbol{w} \in \operatorname{range} T^{k+1}$, 则有 $\boldsymbol{v} \in V$ 使得

$$\boldsymbol{w} = T^{k+1}\boldsymbol{v} = T^k(T\boldsymbol{v}) \in \operatorname{range} T^k.$$

由此得 $\operatorname{range} T^{k+1} \subset \operatorname{range} T^k$. 也就是说,

$$V = \operatorname{range} T^0 \supset \operatorname{range} T^1 \supset \cdots \supset \operatorname{range} T^k \supset \operatorname{range} T^{k+1} \supset \cdots.$$

下面的命题表明, 在上式中, 一旦指数达到了 V 的维数, 包含就变成了相等.

零空间的相应包含关系 (8.4) 与此恰好相反.

8.9 命题: 若 $T \in \mathcal{L}(V)$, 则

$$\operatorname{range} T^{\dim V} = \operatorname{range} T^{\dim V+1} = \operatorname{range} T^{\dim V+2} = \cdots.$$

证明: 我们本可以从头来证明, 但我们更愿意利用已经证明过的关于零空间的相应结果. 设 $m > \dim V$, 则

$$\begin{aligned}\dim \operatorname{range} T^m &= \dim V - \dim \operatorname{null} T^m \\ &= \dim V - \dim \operatorname{null} T^{\dim V} \\ &= \dim \operatorname{range} T^{\dim V},\end{aligned}$$

其中第一个和第三个等式由 3.4 得到, 第二个等式由 8.6 得到. 已

经知道 $\operatorname{range} T^{\dim V} \supset \operatorname{range} T^m$, 刚刚又证明了 $\dim \operatorname{range} T^{\dim V} = \dim \operatorname{range} T^m$, 故 $\operatorname{range} T^{\dim V} = \operatorname{range} T^m$. ■

§8.2 特征多项式

设 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$. 我们知道, V 有基使得 T 关于此基有上三角矩阵 (参见 5.13). 一般来说, 这个矩阵不是唯一的 —— T 可能关于 V 的很多不同的基都具有上三角矩阵, 而这些上三角矩阵可能是不同的. 但是, 这些矩阵的对角线上一定都恰好包含了 T 的所有本征值 (参见 5.18). 因此, 如果 T 有 $\dim V$ 个不同的本征值, 那么每个本征值在 T 的任何上三角矩阵的对角线上一定都恰好出现一次.

如果 T 的互异本征值的个数少于 $\dim V$ 会怎么样呢? 这样的情况常常会出现, 此时每个本征值在 T 的任何上三角矩阵的对角线上一定都至少出现一次, 但是其中必然有一些会重复. 一个取定的本征值重复的次数与 V 的基的选取有关吗?

如果 T 关于某个基恰好具有对角矩阵 $\boldsymbol{A}$, 那么你应该验证, λ 在 $\boldsymbol{A}$ 的对角线上恰好出现 $\dim \operatorname{null}(T-\lambda I)$ 次.

你可能会猜测, 一个数 λ 在 T 的一个上三角矩阵的对角线上恰好出现 $\dim \operatorname{null}\,(T-\lambda I)$ 次. 一般来说, 这是不对的. 例如, 考虑 $\mathbf{C}^2$ 上的一个算子, 它关于标准基的矩阵为上三角矩阵

$$\begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}.$$

对于这个算子有 $\dim \operatorname{null}\,(T-5I) = 1$, 而 5 在对角线上出现两次. 但是要注意, 对于这个算子有 $\dim \operatorname{null}\,(T-5I)^2 = 2$. 这个例子阐明了一种普遍情况 —— 数 λ 在 T 的一个上三角矩阵的对角线上恰好出现 $\dim \operatorname{null}\,(T-\lambda I)^{\dim V}$ 次, 这正是我们在下面的定理中将要证明的. 因为 $\operatorname{null}\,(T-\lambda I)^{\dim V}$ 仅依赖于 T 和 λ, 而不依赖于基的选取, 这说明本征值在 T 的上三角矩阵的对角线上出现的次数和特定的基的选取无关. 这个结果将成为我们分析复向量空间上算子的结构的主要工具.

8.10 定理: 设 $T \in \mathcal{L}(V)$, 并且 $\lambda \in \mathbf{F}$, 则对 V 的每个使得 T 具有上三角矩阵的基, λ 在 T 的矩阵对角线上都恰好出现 $\dim \operatorname{null} (T-\lambda I)^{\dim V}$ 次.

证明: 不失一般性, 设 $\lambda = 0$ (如果在这种情况下证明了定理, 以 $T-\lambda I$ 代替 T 便可以得到一般情况的证明).

为了方便, 令 $n = \dim V$. 我们通过对 n 用归纳法来证明这个定理. 若 $n=1$, 则结果显然成立. 因此假设 $n>1$, 并设结果对维数为 $n-1$ 的空间成立.

设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, 使得 T 关于此基有上三角矩阵

回想一下, 矩阵中星号经常用来表示未知的或者我们不关心的元素.

8.11
$$\begin{bmatrix} \lambda_1 & & & * \\ & \ddots & & \\ & & \lambda_{n-1} & \\ 0 & & & \lambda_n \end{bmatrix}.$$

设 $U = \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1})$. 显然 U 在 T 下是不变的 (参见 5.12), 并且 $T|_U$ 关于基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1})$ 的矩阵为

8.12
$$\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_{n-1} \end{bmatrix}.$$

因此, 由归纳法假设, 0 在 8.12 的对角线上出现 $\dim \operatorname{null} (T|_U)^{n-1}$ 次. 我们知道 $\operatorname{null} (T|_U)^{n-1} = \operatorname{null} (T|_U)^n$ (因为 U 的维数为 $n-1$; 参见 8.6), 于是

8.13 0 在 8.12 的对角线上出现 $\dim \operatorname{null} (T|_U)^n$ 次.

根据 λ_n 是否为 0 分两种情况来证明. 首先考虑 $\lambda_n \neq 0$ 的情况. 往证在这种情况下有

8.14
$$\operatorname{null} T^n \subset U.$$

一旦证明了这一点, 就会知道 $\operatorname{null} T^n = \operatorname{null} (T|_U)^n$, 从而, 由 8.13 知, 0 在 8.11 的对角线上恰好出现 $\dim \operatorname{null} T^n$ 次, 这样就完成 $\lambda_n \neq 0$ 的情形的证明了.

因为 $\mathcal{M}(T)$ 是由 8.11 给出的, 所以

$$\mathcal{M}(T^n) = \mathcal{M}(T)^n = \begin{bmatrix} \lambda_1^n & & & * \\ & \ddots & & \\ & & \lambda_{n-1}^n & \\ 0 & & & \lambda_n^n \end{bmatrix}.$$

即存在 $\boldsymbol{u} \in U$ 使得

$$T^n \boldsymbol{v}_n = u + \lambda_n^n \boldsymbol{v}_n$$

要证明 8.14 (仍然假设 $\lambda_n \neq 0$), 设 $\boldsymbol{v} \in \text{null } T^n$. 我们可以把 $\boldsymbol{v}$ 写成

$$\boldsymbol{v} = \tilde{\boldsymbol{u}} + a\boldsymbol{v}_n,$$

其中 $\tilde{\boldsymbol{u}} \in U$, $a \in \mathbf{F}$. 因此

$$\mathbf{0} = T^n \boldsymbol{v} = T^n \tilde{\boldsymbol{u}} + aT^n \boldsymbol{v}_n = T^n \tilde{\boldsymbol{u}} + a\boldsymbol{u} + a\lambda_n^n \boldsymbol{v}_n.$$

因为 $T^n \tilde{\boldsymbol{u}}$ 和 $a\boldsymbol{u}$ 都包含于 U, 并且 $\boldsymbol{v}_n \notin U$, 所以 $a{\lambda_n}^n = 0$. 再由 $\lambda_n \neq 0$ 可得 $a = 0$. 于是 $\boldsymbol{v} = \tilde{\boldsymbol{u}} \in U$, 这就完成了 8.14 的证明.

现在考虑 $\lambda_n = 0$ 的情况. 在这种情况下, 往证

8.15 $$\dim \text{null } T^n = \dim \text{null } (T|_U)^n + 1,$$

再利用 8.13 就可以完成 $\lambda_n = 0$ 的情形的证明.

利用两个子空间之和的维数公式 (2.18) 可得

$$\begin{aligned} \dim \text{null } T^n &= \dim(U \cap \text{null } T^n) + \dim(U + \text{null } T^n) - \dim U \\ &= \dim \text{null } (T|_U)^n + \dim(U + \text{null } T^n) - (n-1). \end{aligned}$$

假如我们能够证明 null T^n 包含一个不属于 U 的向量, 那么

$$n = \dim V \geqslant \dim(U + \text{null } T^n) > \dim U = n - 1,$$

由此得 $\dim(U + \text{null } T^n) = n$, 再结合上面关于 $\dim \text{null } T^n$ 的公式即得 8.15. 于是, 要完成证明只需证明 null T^n 包含一个不属于 U 的向量.

考虑如何才能找到一个包含于 null T^n 而不包含于 U 的向量. 可以试试如下形式的向量

$$\boldsymbol{u} - \boldsymbol{v}_n,$$

其中 $\boldsymbol{u} \in U$. 至少我们可以保证这样的向量不在 U 中, 能不能

找到 $\boldsymbol{u} \in U$ 使得上式中的向量在 null T^n 中呢? 计算

$$T^n(\boldsymbol{u} - \boldsymbol{v}_n) = T^n\boldsymbol{u} - T^n\boldsymbol{v}_n.$$

为使上面的向量等于$\mathbf{0}$, 必须选取 (如果可能的话) $\boldsymbol{u} \in U$ 使得 $T^n\boldsymbol{u} = T^n\boldsymbol{v}_n$. 如果 $T^n\boldsymbol{v}_n \in \operatorname{range}(T|_U)^n$, 就能够做到这一点. 因为 8.11 是 T 关于 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 的矩阵, 所以 $T\boldsymbol{v}_n \in U$ (回想一下, 我们现在讨论的是 $\lambda_n = 0$ 的情况). 因此

$$T^n\boldsymbol{v}_n = T^{n-1}(T\boldsymbol{v}_n) \in \operatorname{range}(T|_U)^{n-1} = \operatorname{range}(T|_U)^n,$$

其中最后一个等式由 8.9 得到. 也就是说, 确实可以取到 $\boldsymbol{u} \in U$ 使得 $\boldsymbol{u} - \boldsymbol{v}_n \in \operatorname{null} T^n$. ■

设 $T \in \mathcal{L}(V)$. T 的本征值 λ 的**重数 (multiplicity)** 定义为相应于 λ 的广义本征向量所构成的子空间的维数. 也就是说, T 的本征值 λ 的重数等于 $\dim \operatorname{null}\,(T - \lambda I)^{\dim V}$. 如果 T 关于 V 的某个基有上三角矩阵 (在 $\mathbf{F} = \mathbf{C}$ 时总是出现), 那么 λ 的重数就是 λ 在这个矩阵的对角线上出现的次数 (由上一个定理).

我们对重数的定义显然和 T 的几何性质有关. 大部分教材利用由行列式所定义的多项式的根的重数来定义重数. 这两种定义实际上是等价的.

作为重数的例子, 考虑如下定义的 $T \in \mathcal{L}(\mathbf{F}^3)$

8.16 $$T(z_1, z_2, z_3) = (0, z_1, 5z_3).$$

你应该验证, 0 是 T 的重数为 2 的本征值, 5 是 T 的重数为 1 的本征值, 并且 T 没有其他本征值. 又如, 若算子 $T \in \mathcal{L}(\mathbf{F}^3)$ 的矩阵为

8.17 $$\begin{bmatrix} 6 & 7 & 7 \\ 0 & 6 & 7 \\ 0 & 0 & 7 \end{bmatrix},$$

那么 6 是 T 的重数为 2 的本征值, 7 是 T 的重数为 1 的本征值 (由定理 8.10 得).

在上面的每个例子中, T 的本征值的重数之和都等于 3, 即 T 的定义域的维数. 下一个命题表明这在复向量空间上总成立.

8.18 命题： 设 V 是复向量空间，并且 $T \in \mathcal{L}(V)$，那么 T 的所有本征值的重数之和等于 $\dim V$.

证明： 设 V 是复向量空间，并且 $T \in \mathcal{L}(V)$，那么 V 有基使得 T 关于此基有上三角矩阵 (由 5.13). λ 的重数等于 λ 在这个矩阵的对角线上出现的次数 (由 8.10). 因为这个矩阵的对角线长度为 $\dim V$，所以 T 的所有本征值的重数之和一定等于 $\dim V$. ■

设 V 是复向量空间，并且 $T \in \mathcal{L}(V)$. 令 $\lambda_1, \cdots, \lambda_m$ 表示 T 的所有互不相同的本征值. 令 d_j 表示 T 的本征值 λ_j 的重数. 多项式

$$(z-\lambda_1)^{d_1} \cdots (z-\lambda_m)^{d_m}$$

称为 T 的**特征多项式 (characteristic polynomial)**. 注意到，T 的特征多项式的次数等于 $\dim V$ (由 8.18). 显然，T 的特征多项式的根等于 T 的本征值. 例如，8.16 所定义的算子 $T \in \mathcal{L}(\mathbf{C}^3)$ 的特征多项式等于 $z^2(z-5)$.

大部分教材利用行列式来定义特征多项式. 这里所采用的方法更简单，并且由此得到了凯莱–哈密顿定理的一个简单证明.

对于复向量空间上算子的特征多项式还有另一种描述. 设 V 是复向量空间，并且 $T \in \mathcal{L}(V)$. 考虑使得 T 具有上三角矩阵

8.19
$$\mathcal{M}(T) = \begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

的 V 的任意一个基. 由 8.10 立即可得 T 的特征多项式为

$$(z-\lambda_1) \cdots (z-\lambda_n).$$

来看另一个这样的例子，如果 $T \in \mathcal{L}(\mathbf{C}^3)$ 是一个矩阵为 8.17 的算子，那么 T 的特征多项式等于 $(z-6)^2(z-7)$.

下一章我们将定义实向量空间上算子的特征多项式，并且证明下一个结果对实向量空间也成立.

8.20 凯莱–哈密顿定理 (Cayley-Hamilton Theorem): 设 V 是复向量空间, $T \in \mathcal{L}(V)$, 并设 q 表示 T 的特征多项式, 那么 $q(T) = 0$.

英国数学家凯莱 (Arthur Cayley) 于 1842 年完成了他的学士学位, 此前他已经发表了三篇数学论文. 爱尔兰数学家哈密顿 (William Hamilton) 于 1827 年成为教授, 当时他只有 22 岁, 还是一个本科生!

证明: 设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, 使得 T 关于此基有形如 8.19 的上三角矩阵. 要证明 $q(T) = 0$, 只需证明 $q(T)\boldsymbol{v}_j = \boldsymbol{0}$, $j = 1, \cdots, n$. 为此, 只需证明

8.21 $$(T - \lambda_1 I) \cdots (T - \lambda_j I)\boldsymbol{v}_j = \boldsymbol{0}, \quad j = 1, \cdots, n.$$

对 j 用归纳法来证明 8.21. 首先, 设 $j = 1$. 因为 $\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))$ 由 8.19 给出, 所以 $T\boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1$, 即当 $j = 1$ 时 8.21 成立.

现在设 $1 < j \leqslant n$, 并且

$$\begin{aligned}
\boldsymbol{0} &= (T - \lambda_1 I)\boldsymbol{v}_1 \\
&= (T - \lambda_1 I)(T - \lambda_2 I)\boldsymbol{v}_2 \\
&\ \vdots \\
&= (T - \lambda_1 I) \cdots (T - \lambda_{j-1} I)\boldsymbol{v}_{j-1}.
\end{aligned}$$

因为 $\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))$ 由 8.19 给出, 所以

$$(T - \lambda_j I)\boldsymbol{v}_j \in \operatorname{span}(\boldsymbol{v}_1, \cdots \boldsymbol{v}_{j-1}).$$

因此, 由归纳法假设, $(T - \lambda_1 I) \cdots (T - \lambda_{j-1} I)$ 作用在 $(T - \lambda_j I)\boldsymbol{v}_j$ 上得 $\boldsymbol{0}$, 即 8.21 成立. ■

§8.3 算子的分解

前面我们看到, 即使是复向量空间上的算子, 其定义域也未必能分解成一些由算子的本征向量组成的不变子空间的直和. 本节我们将会看到, 每个复向量空间上的算子都有足够多的广义本征向量来给出一个分解.

前面我们观察到, 若 $T \in \mathcal{L}(V)$, 则 null T 在 T 下是不变的. 现在我们要证明, T 的任何多项式的零空间在 T 下也都是不变的.

8.22 命题： 如果 $T \in \mathcal{L}(V)$, 并且 $p \in \mathcal{P}(\mathbf{F})$, 那么 null $p(T)$ 在 T 下是不变的.

证明： 设 $T \in \mathcal{L}(V)$, $p \in \mathcal{P}(F)$, 并设 $\boldsymbol{v} \in \operatorname{null} p(T)$, 则 $p(T)\boldsymbol{v} = \mathbf{0}$. 因此

$$(p(T))(T\boldsymbol{v}) = T(p(T)\boldsymbol{v}) = T(\mathbf{0}) = \mathbf{0},$$

从而 $T\boldsymbol{v} \in \operatorname{null} p(T)$. 于是 null $p(T)$ 在 T 下是不变的. ■

下面的主要结构定理说明, 复向量空间上的每个算子都可以看成是由几部分组成的, 其中每一部分都是一个幂零算子加上恒等算子的一个标量倍. 事实上, 我们已经完成了所有困难的工作, 所以现在的证明就简单了.

8.23 定理： 假设 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$. 设 $\lambda_1, \cdots, \lambda_m$ 是 T 的不同本征值, 并且 $U_1, \cdots, U_m$ 分别是相应的广义本征向量的子空间, 那么

(a) $V = U_1 \oplus \cdots \oplus U_m$;

(b) 每个 U_j 在 T 下都是不变的;

(c) 每个 $(T - \lambda_j I)|_{U_j}$ 都是幂零的.

证明： 注意到对每个 j 都有 $U_j = \operatorname{null} (T - \lambda_j I)^{\dim V}$ (由 8.7). 由 8.22 (取 $p(z) = (z - \lambda_j)^{\dim V}$) 可得 (b). 显然, 由定义可得 (c).

要证明 (a), 回想一下, λ_j 作为 T 的本征值的重数定义为 $\dim U_j$. 这些重数之和等于 $\dim V$ (参见 8.18); 因此

8.24 $$\dim V = \dim U_1 + \cdots + \dim U_m.$$

令 $U = U_1 + \cdots + U_m$. 显然 U 在 T 下是不变的. 于是可以定义 $S \in \mathcal{L}(U)$ 为

$$S = T|_U.$$

注意到 S 与 T 具有相同的本征值, 并且重数相同, 这是因为 T 的所有广义本征向量都在 S 的定义域 U 中. 于是, 对 S 应用 8.18 可得

$$\dim U = \dim U_1 + \cdots + \dim U_m.$$

再结合 8.24 可得 $\dim V = \dim U$. 因为 U 是 V 的子空间, 所以 $V = U$. 也就是说

$$V = U_1 + \cdots + U_m.$$

再结合 8.24, 并利用 2.19 即可得 (a). ■

我们知道, 复向量空间上的算子可能没有足够多的本征向量来组成定义域的基. 下一个结果表明复向量空间上的算子有足够多的广义本征向量来组成定义域的基.

8.25 推论: 设 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$, 那么 V 有一个由 T 的广义本征向量组成的基.

证明: 给 8.23 中的每个 U_j 取一个基. 将所有这些基放在一起就得到 V 的一个由 T 的广义本征向量组成的基. ■

给定 V 上的算子 T, 我们想要找到 V 的一个基, 使得 T 在此基下的矩阵尽可能简单, 即这个矩阵包含很多 0. 首先来证明, 如果 N 是幂零的, 那么可以取 V 的一个基, 使得 N 关于此基的矩阵有一半以上的元素都等于 0.

8.26 引理: 设 N 是 V 上的幂零算子, 那么 V 有一个基使得 N 关于此基的矩阵形如

如果 V 是复向量空间, 那么这个引理的证明容易由本章的习题 6、5.13 和 5.18 得到. 但是这里的证明思想要比 5.13 的证明思想简单, 并且对实向量空间和复向量空间都适用.

8.27
$$\begin{bmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{bmatrix};$$

这里对角线上和对角线下方的元素都是 0.

证明: 首先取 $\operatorname{null} N$ 的一个基, 将它扩充成 $\operatorname{null} N^2$ 的基, 再扩充成 $\operatorname{null} N^3$ 的基. 如此下去, 最终得到 V 的一个基 (因为当 m 足够大时有 $\operatorname{null} N^m = V$).

现在我们来考虑 N 在这个基下的矩阵. 第一列, 或许前几列, 全部由 0 组成, 因为相应的基向量包含于 null N. 下一组列来自 null N^2 中的基向量. 任意这样的向量被 N 作用后都是 null N 中的向量; 也就是说, 所得到的向量是前面的基向量的线性组合. 因此这些列中所有的非零元素一定都出现在对角线的上方. 再下一组列来自 null N^3 的基向量. 任意这样的向量被 N 作用都是 null N^2 中的向量; 也就是说, 所得到的向量是前面的基向量的线性组合. 于是, 这又表明这些列中的非零元素一定都出现在对角线上方. 如此继续下去即可完成证明. ■

注意, 在下一个定理中, T 的矩阵比上三角形式的矩阵有更多的零.

8.28 定理: 假设 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$. 设 $\lambda_1, \cdots, \lambda_m$ 是 T 的所有互不相同的本征值, 那么 V 有一个基使得 T 关于此基有如下的分块对角矩阵

$$\begin{bmatrix} \boldsymbol{A}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中每个 $\boldsymbol{A}_j$ 都是如下形式的上三角矩阵

8.29
$$\boldsymbol{A}_j = \begin{bmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{bmatrix}.$$

证明: 对 $j = 1, \cdots, m$, 设 U_j 表示由 T 的相应于 λ_j 的广义本征向量构成的子空间, 则 $(T - \lambda_j I)|_{U_j}$ 是幂零的 (参见 8.23(c)). 对每个 j, 取 U_j 的一个基使得 $(T - \lambda_j I)|_{U_j}$ 关于此基的矩阵形如 8.27. 于是, $T|_{U_j}$ 关于此基的矩阵形如 8.29. 将这些 U_j 的基放在一起就组成 V 的一个基 (由 8.23(a)), 而且 T 关于此基的矩阵就具有想要的形式. ■

§8.4 平 方 根

回想一下, 算子 $T \in \mathcal{L}(V)$ 的平方根是一个算子, $S \in \mathcal{L}(V)$, 满足 $S^2 = T$. 作为上一节中主要结构定理 (8.23) 的应用, 本节将证明复向量空间上的每个可逆算子都有平方根.

每个复数都有平方根, 但复向量空间上的算子并不都有平方根. 本章习题 4 给出了 $\mathbf{C}^3$ 上的一个没有平方根的算子的例子. 我们很快就会看到, 这个特殊算子的不可逆性并不是偶然的. 首先证明, 恒等算子加上一个幂零算子总有平方根.

8.30 引理: *设 $N \in \mathcal{L}(V)$ 是幂零的, 则 $I + N$ 有平方根.*

证明: 考虑函数 $\sqrt{1+x}$ 的泰勒级数

8.31
$$\sqrt{1+x} = 1 + a_1x + a_2x^2 + \cdots .$$

我们无需找出所有系数的显式公式, 也不用担心这个无限和是否收敛, 因为我们只想从中受到启发, 而不是要把它作为正式证明的一部分.

因为 $a_1 = 1/2$, 所以这个公式表明, 当 x 很小时, $1 + x/2$ 是 $\sqrt{1+x}$ 的一个很好的估计.

因为 N 是幂零的, 所以存在正整数 m 使得 $N^m = 0$. 在 8.31 中, 如果用 N 替换 x, 用 I 替换 1, 那么右端的无限和就成为一个有限和 (因为对所有的 $j \geqslant m$ 都有 $N^j = 0$). 也就是说, 可以猜测 $I + N$ 有如下形式的平方根

$$I + a_1N + a_2N^2 + \cdots + a_{m-1}N^{m-1}.$$

根据这个猜测, 试取 $a_1, a_2, \cdots, a_{m-1}$ 使得上面算子的平方等于 $I + N$. 现在

$$\begin{aligned}(I + a_1N + a_2N^2 + a_3N^3 + \cdots + a_{m-1}N^{m-1})^2 \\ = I + 2a_1N + (2a_2 + a_1{}^2)N^2 + (2a_3 + 2a_1a_2)N^3 + \cdots \\ + (2a_{m-1} + \text{包含 } a_1, \cdots, a_{m-2} \text{ 的项})N^{m-1}.\end{aligned}$$

我们希望上式的右端等于 $I + N$. 于是, 取 a_1 使得 $2a_1 = 1$ (从而 $a_1 = 1/2$). 再取 a_2 使得 $2a_2 + a_1^2 = 0$ (从而 $a_2 = -1/8$). 然后再取 a_3 使得上 式右端 N^3 的系数等于 0 (从而 $a_3 = 1/16$).

对 $j=4,\cdots,m-1$ 如此进行下去, 每一步都解出一个 a_j 使得上式右端 N^j 的系数等于 0. 事实上, 我们并不关心这些 a_j 的显式公式, 只需知道这些 a_j 的选取能够使得 $I+N$ 有平方根. ■

上一个引理对实向量空间和复向量空间都成立, 但下一个结果却只对复向量空间成立.

实向量空间上的可逆算子未必有平方根. 例如, **R** 上乘以 -1 的算子没有平方根, 因为任何实数的平方都不等于 -1.

8.32 定理: *设 V 是复向量空间. 如果 $T\in\mathcal{L}(V)$ 是可逆的, 那么 T 有平方根.*

证明: 设 $T\in\mathcal{L}(V)$ 是可逆的, 并设 $\lambda_1,\cdots,\lambda_m$ 是 T 的所有互不相同的本征值, $U_1,\cdots,U_m$ 分别是相应的广义本征向量所构成的子空间. 对每个 j 都存在一个幂零算子 $N_j\in\mathcal{L}(U_j)$, 使得 $T|_{U_j}=\lambda_j I+N_j$ (参见 8.23(c)). 因为 T 是可逆的, 所以每个 λ_j 都不等于 0, 于是对每个 j 都有

$$T|_{U_j}=\lambda_j\left(I+\frac{N_j}{\lambda_j}\right).$$

显然 N_j/λ_j 是幂零的, 于是 $I+N_j/\lambda_j$ 有平方根 (由 8.30). 用复数 λ_j 的一个平方根乘以 $I+N_j/\lambda_j$ 的一个平方根就得到 $T|_{U_j}$ 的一个平方根 S_j.

一个向量 $\boldsymbol{v}\in V$ 可以唯一地写成如下形式

$$\boldsymbol{v}=\boldsymbol{u}_1+\cdots+\boldsymbol{u}_m,$$

其中每个 $\boldsymbol{u}_j\in U_j$ (参见 8.23). 利用这个分解, 定义算子 $S\in\mathcal{L}(V)$ 如下

$$S\boldsymbol{v}=S_1\boldsymbol{u}_1+\cdots+S_m\boldsymbol{u}_m.$$

你应该验证, 这个算子 S 就是 T 的平方根. ■

仿照本节的方法, 你应该能够证明: 如果 V 是复向量空间, 并且 $T\in\mathcal{L}(V)$ 是可逆的, 那么对每个正整数 k, T 都有 k 次根.

§8.5 极小多项式

我们即将看到, 给定有限维向量空间上的一个算子, 存在唯一作用在这个算子上等于 0 的次数最小的首一多项式. 这个多项式称为该算子的极小多项式. 极小多项式是本节关注的焦点.

首一多项式 (monic polynomial) 是指最高次数的项的系数为 1 的多项式. 例如, $2+3z^2+z^8$ 是首一多项式.

设 $T \in \mathcal{L}(V)$, 其中 $\dim V = n$, 那么

$$(I, T, T^2, \cdots, T^{n^2})$$

在 $\mathcal{L}(V)$ 中不可能是线性无关的, 这是因为 $\mathcal{L}(V)$ 的维数是 n^2 (参见 3.20), 而这里有 n^2+1 个算子. 设 m 是使得

8.33 $$(I, T, T^2, \cdots, T^m)$$

线性相关的最小的正整数. 线性相关性引理 (2.4) 表明在上述组中有一个算子是它前面的那些算子的线性组合. 因为 m 被取成是使得 8.33 线性相关的最小正整数, 所以 T^m 是 $(I, T, T^2, \cdots, T^{m-1})$ 的线性组合. 于是有标量 $a_0, a_1, a_2, \cdots, a_{m-1} \in \mathbf{F}$ 使得

$$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m = 0.$$

标量 $a_0, a_1, a_2, \cdots, a_{m-1} \in \mathbf{F}$ 的选取是唯一的, 因为如果有两种不同的选取就与 m 的取法矛盾 (将上述这种形式的两个不同的等式相减, 将得到一个比 8.33 更短的线性无关组). 多项式

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

称为 T 的**极小多项式 (minimal polynomial)**. 它是使得 $p(T) = 0$ 的次数最小的首一多项式 $p \in \mathcal{P}(\mathbf{F})$.

例如, 恒等算子 I 的极小多项式是 $z-1$. 你应该验证, $\mathbf{F}^2$ 上具有矩阵 $\left[\begin{smallmatrix} 4 & 1 \\ 0 & 5 \end{smallmatrix}\right]$ 的算子的极小多项式是 $20 - 9z + z^2$.

显然 V 上每个算子的极小多项式的次数都不超过 $(\dim V)^2$. 凯莱–哈密顿定理 (8.20) 告诉我们, 如果 V 是复向量空间, 那么 V 上每个算子的极小多项式的次数最多为 $\dim V$. 下一章我们将看到这一显著的改进对实向量空间也成立.

称多项式 $p \in \mathcal{P}(\mathbf{F})$ **整除 (divide)** 多项式 $q \in \mathcal{P}(\mathbf{F})$, 如果存在一个多项式 $s \in \mathcal{P}(\mathbf{F})$ 使得 $q = sp$. 也就是说, p 整除 q 当且仅当 4.6 中的余项 r 可以取成 0. 例如, 多项式 $(1+3z)^2$ 整除 $5+32z+57z^2+18z^3$, 因为 $5+32z+57z^2+18z^3 = (2z+5)(1+3z)^2$. 显然, 非零常数多项式整除任意多项式.

注意, 由 4.1 立即可得, $(z-\lambda)$ 整除多项式 q 当且仅当 λ 是 q 的根.

下一个结果完全刻画了作用在一个算子上等于 0 的多项式.

8.34 定理: 设 $T \in \mathcal{L}(V)$, $q \in \mathcal{P}(\mathbf{F})$, 那么 $q(T) = 0$ 当且仅当 T 的极小多项式整除 q.

证明: 设 p 是 T 的极小多项式.

首先证明容易的一方面. 设 p 整除 q, 那么存在一个多项式 $s \in \mathcal{P}(\mathbf{F})$ 使得 $q = sp$. 于是

$$q(T) = s(T)p(T) = s(T)0 = 0.$$

要证明另一个方面, 设 $q(T) = 0$. 由带余除法 (4.5), 存在多项式 $s, r \in \mathcal{P}(\mathbf{F})$ 使得

8.35
$$q = sp + r$$

并且 $\deg r < \deg p$. 于是

$$0 = q(T) = s(T)p(T) + r(T) = r(T).$$

因为 p 是 T 的极小多项式, 并且 $\deg r < \deg p$, 所以由上式可得 $r = 0$. 因此 8.35 成为 $q = sp$, 即 p 整除 q. ■

现在利用一个算子的极小多项式来描述它的本征值.

8.36 定理: 设 $T \in \mathcal{L}(V)$, 那么 T 的极小多项式的根恰好是 T 的本征值.

证明：设

$$p(z)=a_0+a_1z+a_2z^2+\cdots+a_{m-1}z^{m-1}+z^m$$

是 T 的极小多项式.

首先, 设 $\lambda\in\mathbf{F}$ 是 p 的一个根, 那么 T 的极小多项式可以写成如下形式

$$p(z)=(z-\lambda)q(z),$$

其中 q 是系数在 $\mathbf{F}$ 中的首一多项式 (参见 4.1). 因为 $p(T)=0$, 所以对所有 $\boldsymbol{v}\in V$ 都有

$$\mathbf{0}=(T-\lambda I)(q(T)\boldsymbol{v}).$$

又因为 q 的次数小于极小多项式 p 的次数, 所以至少存在一个向量 $\boldsymbol{v}\in V$ 使得 $q(T)\boldsymbol{v}\neq\mathbf{0}$. 于是由上面的等式可知, λ 是 T 的本征值.

要证明另一个方面, 假设 $\lambda\in\mathbf{F}$ 是 T 的本征值. 设 $\boldsymbol{v}$ 是 V 中的一个使得 $T\boldsymbol{v}=\lambda\boldsymbol{v}$ 的非零向量. 用 T 反复作用等式两端可知, 对每个非负整数 j 都有 $T^j\boldsymbol{v}=\lambda^j\boldsymbol{v}$. 因此

$$\begin{aligned}\mathbf{0}=p(T)\boldsymbol{v}&=(a_0+a_1T+a_2T^2+\cdots+a_{m-1}T^{m-1}+T^m)\boldsymbol{v}\\&=(a_0+a_1\lambda+a_2\lambda^2+\cdots+a_{m-1}\lambda^{m-1}+\lambda^m)\boldsymbol{v}\\&=p(\lambda)\boldsymbol{v}\end{aligned}$$

因为 $\boldsymbol{v}\neq\mathbf{0}$, 所以由上面的等式得 $p(\lambda)=0$. ■

假设已知某个算子 $T\in\mathcal{L}(V)$ (关于某个基) 的矩阵的具体形式. 为求 T 的极小多项式, 对 $m=1,2,\cdots$, 考虑

$$(\mathcal{M}(I),\mathcal{M}(T),\mathcal{M}(T)^2,\cdots,\mathcal{M}(T)^m),$$

直到这个组是线性相关的. 然后求标量 $a_0,a_1,a_2,\cdots,a_{m-1}\in\mathbf{F}$ 使得

$$a_0\mathcal{M}(I)+a_1\mathcal{M}(T)+a_2\mathcal{M}(T)^2+\cdots+a_{m-1}\mathcal{M}(T)^{m-1}+\mathcal{M}(T)^m=0.$$

于是标量 $a_0,a_1,a_2,\cdots,a_{m-1},1$ 就是 T 的极小多项式的系数. 所有这些都可以利用像 Gauss 消元法这种熟悉的方法来计算.

你可以把它看作是关于 m 个变量 $a_0,a_1,\cdots,a_{m-1}$ 的 $(\dim V)^2$ 个方程组成的方程组.

例如, 考虑 $\mathbf{C}^5$ 中具有如下矩阵的算子 T,

8.37
$$\begin{bmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

因为这个矩阵有大量的 0, 所以这里不需要 Gauss 消元法. 只需算出 $\mathcal{M}(T)$ 的幂, 并注意到只有到 5 次幂时它们才是线性相关的. 通过计算可知, T 的极小多项式等于

8.38
$$z^5 - 6z + 3$$

现在, 这个给定算子的本征值是什么呢? 由 8.36 知 T 的本征值等于下面方程的解

$$z^5 - 6z + 3 = 0.$$

可惜这个方程的解不能用有理数、有理数的任意次根以及通常的四则运算来计算 (其证明远远超出了线性代数的范围). 因此无法用我们所熟悉的形式给出 T 的任何本征值的精确表达, 但是数值方法可以给出 T 的很好的近似本征值, 比如

$$-1.67 \quad 0.51, \quad 1.40, \quad -0.12 + 1.59i, \quad -0.12 - 1.59i.$$

我们在这里不讨论数值方法. 注意到, 就像实系数多项式的非实数根是成对出现的一样 (参见 4.10), 非实数本征值与其复共轭也是成对出现的.

设 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$. 由凯莱–哈密顿定理 (8.20) 和 8.34 可知, T 的极小多项式整除 T 的特征多项式. 这两个多项式都是首一的. 因此, 如果 T 的极小多项式的次数为 $\dim V$, 那么它一定等于 T 的特征多项式的次数. 例如, 若 $\mathbf{C}^5$ 上算子 T 的矩阵为 8.37, 则 T 的特征多项式和 T 的极小多项式都等于 8.38.

§8.6 约 当 形

我们知道, 如果 V 是复向量空间, 那么对于每个 $T \in \mathcal{L}(V)$, V 都有一个基使得 T 关于此基有较好形式的上三角矩阵 (参见 8.28). 本节将会得到更好的结论 —— V 有一个基, 使得 T 关于此基的矩阵除了对角线上以及紧位于对角线上方的直线上的元素之外, 其余元素都为零.

首先来描述幂零算子. 例如, 考虑幂零算子 $N \in \mathcal{L}(\mathbf{F}^n)$,

$$N(z_1, \cdots, z_n) = (0, z_1, \cdots, z_{n-1}).$$

如果 $\boldsymbol{v} = (1, 0, \cdots, 0)$, 那么显然 $(\boldsymbol{v}, N\boldsymbol{v}, \cdots, N^{n-1}\boldsymbol{v})$ 是 $\mathbf{F}^n$ 的基, 并且 $(N^{n-1}\boldsymbol{v})$ 是 1 维子空间 null N 的基.

又如, 考虑幂零算子 $N \in \mathcal{L}(\mathbf{F}^5)$,

8.39 $$N(z_1, z_2, z_3, z_4, z_5) = (0, z_1, z_2, 0, z_4).$$

与上一段所讨论的幂零算子不同, 对于这个幂零算子没有向量 $\boldsymbol{v} \in \mathbf{F}^5$ 能使 $(\boldsymbol{v}, N\boldsymbol{v}, N^2\boldsymbol{v}, N^3\boldsymbol{v}, N^4\boldsymbol{v})$ 是 $\mathbf{F}^5$ 的基. 但是, 如果 $\boldsymbol{v}_1 = (1, 0, 0, 0, 0)$, $\boldsymbol{v}_2 = (0, 0, 0, 1, 0)$, 那么 $(\boldsymbol{v}_1, N\boldsymbol{v}_1, N^2\boldsymbol{v}_1, \boldsymbol{v}_2, N\boldsymbol{v}_2)$ 是 $\mathbf{F}^5$ 的基, 并且 $(N^2\boldsymbol{v}_1, N\boldsymbol{v}_2)$ 是 2 维子空间 null N 的基.

设 $N \in \mathcal{L}(V)$ 是幂零的. 对每个非零向量 $\boldsymbol{v} \in V$, 令 $m(\boldsymbol{v})$ 表示使得 $N^{m(\boldsymbol{v})}\boldsymbol{v} \neq \mathbf{0}$ 的最大非负整数. 例如, 若 $N \in \mathcal{L}(\mathbf{F}^5)$ 是由 8.39 定义的, 那么 $m(1, 0, 0, 0, 0) = 2$.

显然 $m(\boldsymbol{v})$ 依赖于 N 和 $\boldsymbol{v}$, 但 N 的选取在上下文中是明显的.

下面的引理表明, 每个幂零算子 $N \in \mathcal{L}(V)$ 都与 8.39 所定义的例子性质相似, 这是因为存在有限个向量 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \in V$, 使得形如 $N^j\boldsymbol{v}_r$ 的非零向量组成 V 的一个基; 这里 r 从 1 取到 k, 而 j 从 0 取到 $m(\boldsymbol{v}_r)$.

8.40 引理: 如果 $N \in \mathcal{L}(V)$ 是幂零的, 那么存在向量 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \in V$ 使得

(a) $(\boldsymbol{v}_1, N\boldsymbol{v}_1, \cdots, N^{m(\boldsymbol{v}_1)}\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k, N\boldsymbol{v}_k, \cdots, N^{m(\boldsymbol{v}_k)}\boldsymbol{v}_k)$ 是 V 的基;

(b) $(N^{m(\boldsymbol{v}_1)}\boldsymbol{v}_1, \cdots, N^{m(\boldsymbol{v}_k)}\boldsymbol{v}_k)$ 是 null N 的基.

证明: 设 N 是幂零的, 则 N 不是单的, 于是 $\dim \operatorname{range} N < \dim V$ (参见 3.21). 通过对 V 的维数用归纳法, 可以假设引理对所有维数更小的向量空间都成立. 用 $\operatorname{range} N$ 代替 V, 并且用 $N|_{\operatorname{range} N}$ 代替 N, 于是得到向量 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j \in \operatorname{range} N$, 使得

(i) $(\boldsymbol{u}_1, N\boldsymbol{u}_1, \cdots, N^{m(\boldsymbol{u}_1)}\boldsymbol{u}_1, \cdots, \boldsymbol{u}_j, N\boldsymbol{u}_j, \cdots, N^{m(\boldsymbol{u}_j)}\boldsymbol{u}_j)$ 是 $\operatorname{range} N$ 的基;

(ii) $(N^{m(\boldsymbol{u}_1)}\boldsymbol{u}_1, \cdots, N^{m(\boldsymbol{u}_j)}\boldsymbol{u}_j)$ 是 $\operatorname{null} N \cap \operatorname{range} N$ 的基.

因为每个 $\boldsymbol{u}_r \in \operatorname{range} N$, 所以可取 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_j \in V$ 使得对每个 r 都有 $N\boldsymbol{v}_r = \boldsymbol{u}_r$. 注意到, 对每个 r 都有 $m(\boldsymbol{v}_r) = m(\boldsymbol{u}_r) + 1$.

由 2.13 可得具有这种性质的子空间 W 的存在性.

设 W 是 $\operatorname{null} N$ 的子空间, 使得

8.41
$$\operatorname{null} N = (\operatorname{null} N \cap \operatorname{range} N) \oplus W.$$

取 W 的一个基, 记为 $(\boldsymbol{v}_{j+1}, \cdots, \boldsymbol{v}_k)$. 因为 $\boldsymbol{v}_{j+1}, \cdots, \boldsymbol{v}_k \in \operatorname{null} N$, 所以 $m(\boldsymbol{v}_{j+1}) = \cdots = m(\boldsymbol{v}_k) = 0$.

因为已经构造了 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$, 所以现在需要证明 (a) 和 (b) 成立. 首先证明 (a) 中所谓的基是线性无关的. 为此, 假设

8.42
$$0 = \sum_{r=1}^{k} \sum_{s=0}^{m(\boldsymbol{v}_r)} a_{r,s} N^s(\boldsymbol{v}_r),$$

其中每个 $a_{r,s} \in \mathbf{F}$. 用 N 作用上式两端可得

$$\begin{aligned} 0 &= \sum_{r=1}^{k} \sum_{s=0}^{m(\boldsymbol{v}_r)} a_{r,s} N^{s+1}(\boldsymbol{v}_r) \\ &= \sum_{r=1}^{j} \sum_{s=0}^{m(\boldsymbol{u}_r)} a_{r,s} N^s(\boldsymbol{u}_r). \end{aligned}$$

由上式和 (i) 得 $a_{r,s} = 0$, $1 \leqslant r \leqslant j$, $0 \leqslant s \leqslant m(\boldsymbol{v}_r) - 1$. 因此, 8.42 化简为下面的等式

$$\begin{aligned} 0 = a_{1,m(\boldsymbol{v}_1)} N^{m(\boldsymbol{v}_1)} \boldsymbol{v}_1 + \cdots + a_{j,m(\boldsymbol{v}_j)} N^{m(\boldsymbol{v}_j)} \boldsymbol{v}_j \\ + a_{j+1,0} \boldsymbol{v}_{j+1} + \cdots + a_{k,0} \boldsymbol{v}_k. \end{aligned}$$

第一行右端的项都包含于 $\operatorname{null} N \cap \operatorname{range} N$, 第二行的项都包含

于 W. 因此, 由上式和 8.41 可得

$$\begin{aligned}0&=a_{1,m(\boldsymbol{v}_1)}N^{m(\boldsymbol{v}_1)}\boldsymbol{v}_1+\cdots+a_{j,m(\boldsymbol{v}_j)}N^{m(\boldsymbol{v}_j)}\boldsymbol{v}_j\\ &=a_{1,m(\boldsymbol{v}_1)}N^{m(\boldsymbol{u}_1)}\boldsymbol{u}_1+\cdots+a_{j,m(\boldsymbol{v}_j)}N^{m(\boldsymbol{u}_j)}\boldsymbol{u}_j\end{aligned}\tag{8.43}$$

以及

$$0=a_{j+1,0}\boldsymbol{v}_{j+1}+\cdots+a_{k,0}\boldsymbol{v}_k.\tag{8.44}$$

现在, 由 8.43 和 (ii) 可得 $a_{1,m(\boldsymbol{v}_1)}=\cdots=a_{j,m(\boldsymbol{v}_j)}=0$. 因为 $(\boldsymbol{v}_{j+1},\cdots,\boldsymbol{v}_k)$ 是 W 的基, 所以由 8.44 可得 $a_{j+1,0}=\cdots=a_{k,0}=0$. 从而, 所有的 a 都等于 0, 于是 (a) 中的向量组是线性无关的.

显然, 由 (ii) 可得 $\dim(\operatorname{null} N\cap\operatorname{range} N)=j$. 再结合 8.41 得

$$\dim\operatorname{null} N=k.\tag{8.45}$$

显然, 由 (i) 可得

$$\begin{aligned}\dim\operatorname{range} N&=\sum_{r=1}^{j}(m(\boldsymbol{u}_r)+1)\\ &=\sum_{r=1}^{j}m(\boldsymbol{v}_r).\end{aligned}\tag{8.46}$$

(a) 中向量组的长度为

$$\begin{aligned}\sum_{r=1}^{k}(m(\boldsymbol{v}_r)+1)&=k+\sum_{r=1}^{j}m(\boldsymbol{v}_r)\\ &=\dim\operatorname{null} N+\dim\operatorname{range} N\\ &=\dim V,\end{aligned}$$

其中第二个等式由 8.45 和 8.46 得到, 第三个等式由 3.4 得到. 上式说明 (a) 中向量组的长度为 $\dim V$; 因为这个组是线性无关的, 所以它是 V 的基 (参见 2.17), 这就完成了 (a) 的证明.

最后, 注意到

$$\begin{aligned}&(N^{m(\boldsymbol{v}_1)}\boldsymbol{v}_1,\cdots,N^{m(\boldsymbol{v}_k)}\boldsymbol{v}_k)\\ &=(N^{m(\boldsymbol{u}_1)}\boldsymbol{u}_1,\cdots,N^{m(\boldsymbol{u}_j)}\boldsymbol{u}_j,\boldsymbol{v}_{j+1},\cdots,\boldsymbol{v}_k).\end{aligned}$$

现在, (ii) 和 8.41 表明上面的组是 $\operatorname{null} N$ 的基, 这就证明了 (b).

■

设 $T \in \mathcal{L}(V)$. V 的基称为 T 的**约当基 (Jordan basis)**, 如果 T 关于这个基具有分块对角矩阵

$$\begin{bmatrix} \boldsymbol{A}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中每个 $\boldsymbol{A}_j$ 都是形如

$$\boldsymbol{A}_j = \begin{bmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{bmatrix}.$$

的上三角矩阵. 每个 $\boldsymbol{A}_j$ 的对角线上都是 T 的一些本征值 λ_j, 而紧位于对角线上方的直线上的元素都是 1, 并且所有其他元素都为 0 ($\boldsymbol{A}_j$ 可以只是某个本征值的 1×1 块).

要理解为什么每个 λ_j 都一定是 T 的本征值, 参见 5.18.

因为实向量空间上的算子未必有本征值, 所以也未必有相应的约当基. 于是, 虽然前面的引理对实向量空间和复向量空间都成立, 但是下一个结果需要假设 V 是复向量空间.

法国数学家约当 (Camille Jordan)在 1870 年首先发表了这个定理的证明.

8.47 定理: 设 V 是复向量空间. 如果 $T \in \mathcal{L}(V)$, 那么 V 有一个基是 T 的约当基.

证明: 首先考虑幂零算子 $N \in \mathcal{L}(V)$ 和 8.40 所给出的向量 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \in V$. 注意到, 对于每个 j, N 都将组 $(N^{m(\boldsymbol{v}_j)}\boldsymbol{v}_j, \cdots, N\boldsymbol{v}_j, \boldsymbol{v}_j)$ 中的第一个向量映成 $\mathbf{0}$, 并且 N 将这个组中除第一个向量之外的其余向量映成了它的前一个向量. 也就是说, 如果我们把 8.40(a) 所给出的基顺序反过来的话, 那么就得到 V 的一个基使得 N 具有分块对角矩阵, 其中对角线上的每个矩阵都形如

$$\begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix}.$$

因此定理对幂零算子成立.

现在假设 $T \in \mathcal{L}(V)$, 并设 $\lambda_1, \cdots, \lambda_m$ 是 T 的所有不同本征值, $U_1, \cdots, U_m$ 分别是相应的广义本征向量所构成的子空间. 于是

$$V = U_1 \oplus \cdots \oplus U_m,$$

其中每个 $(T - \lambda_j I)|_{U_j}$ 都是幂零的 (参见 8.23). 由上一段可知, 每个 U_j 都有一个基是 $(T - \lambda_j I)|_{U_j}$ 的约当基. 将这些基放到一起就得到了 V 的一个基, 并且是 T 的约当基. ■

习 题

1. 定义 $T \in \mathcal{L}(\mathbf{C}^2)$ 为

$$T(w, z) = (z, 0).$$

求 T 的所有广义本征向量.

2. 定义 $T \in \mathcal{L}(\mathbf{C}^2)$ 为

$$T(w, z) = (-z, w).$$

求 T 的所有广义本征向量.

3. 设 $T \in \mathcal{L}(V)$, m 是一个正整数, 并且 $\boldsymbol{v} \in V$ 使得 $T^{m-1}\boldsymbol{v} \neq \mathbf{0}$, 但是 $T^m\boldsymbol{v} = \mathbf{0}$. 证明

$$(\boldsymbol{v}, T\boldsymbol{v}, T^2\boldsymbol{v}, \cdots, T^{m-1}\boldsymbol{v})$$

是线性无关的.

4. 设 $T \in \mathcal{L}(\mathbf{C}^3)$ 的定义为 $T(z_1, z_2, z_3) = (z_2, z_3, 0)$. 证明 T 没有平方根. 更确切地说, 证明不存在 $S \in \mathcal{L}(\mathbf{C}^3)$ 使得 $S^2 = T$.

5. 设 $S, T \in \mathcal{L}(V)$. 证明: 如果 ST 是幂零的, 那么 TS 是幂零的.

6. 设 $N \in \mathcal{L}(V)$ 是幂零的. (不用 8.26) 证明 0 是 N 的唯一本征值.

7. 设 V 是内积空间. 证明: 如果 $N \in \mathcal{L}(V)$ 是自伴的并且是幂零的, 那么 $N = 0$.

8. 设 $N \in \mathcal{L}(V)$ 使得 $\operatorname{null} N^{\dim V - 1} \neq \operatorname{null} N^{\dim V}$. 证明: N 是幂零的, 并且对每个使得 $0 \leqslant j \leqslant \dim V$ 的整数 j 都有

$$\dim \operatorname{null} N^j = j.$$

9. 设 $T \in \mathcal{L}(V)$, 并且 m 是一个非负整数使得

$$\operatorname{range} T^m = \operatorname{range} T^{m+1}.$$

证明对每个 $k > m$ 都有 $\operatorname{range} T^k = \operatorname{range} T^m$.

10. 证明或举反例：如果 $T \in \mathcal{L}(V)$, 那么

$$V = \operatorname{null} T \oplus \operatorname{range} T.$$

11. 证明：如果 $T \in \mathcal{L}(V)$, 那么

$$V = \operatorname{null} T^n \oplus \operatorname{range} T^n,$$

其中 $n = \dim V$.

12. 设 V 是复向量空间, $N \in \mathcal{L}(V)$, 并且 0 是 N 的唯一本征值. 证明 N 是幂零的. 举例说明这个结果对实向量空间未必成立.

13. 设 V 是复向量空间, $\dim V = n$, 并且 $T \in \mathcal{L}(V)$ 使得

$$\operatorname{null} T^{n-2} \neq \operatorname{null} T^{n-1}.$$

证明 T 最多有两个不同的本征值.

14. 举例说明 $\mathbf{C}^4$ 上存在特征多项式等于 $(z-7)^2(z-8)^2$ 的算子.

15. 设 V 是复向量空间, $T \in \mathcal{L}(V)$ 使得 5 和 6 是 T 仅有的本征值. 证明

$$(T-5I)^{n-1}(T-6I)^{n-1} = 0,$$

其中 $n = \dim V$.

16. 设 V 是复向量空间, $T \in \mathcal{L}(V)$. 证明 V 有一个由 T 的本征向量组成的基当且仅当 T 的每个广义本征向量都是 T 的本征向量.

对于复向量空间, 这个习题给 5.21 的等价条件列表又添加了一个等价条件.

17. 设 V 是内积空间, 并且 $N \in \mathcal{L}(V)$ 是幂零的. 证明存在 V 的一个规范正交基使得 N 关于此基有上三角矩阵.

18. 定义 $N \in \mathcal{L}(\mathbf{F}^5)$ 为

$$N(x_1, x_2, x_3, x_4, x_5) = (2x_2, 3x_3, -x_4, 4x_5, 0).$$

求 $I + N$ 的一个平方根.

19. 证明：如果 V 是复向量空间，那么 V 上的每个可逆算子都有三次方根.

20. 设 $T \in \mathcal{L}(V)$ 是可逆的. 证明存在一个多项式 $p \in \mathcal{P}(\mathbf{F})$ 使得 $T^{-1} = p(T)$.

21. 举例说明 $\mathbf{C}^3$ 上存在极小多项式等于 z^2 的算子.

22. 举例说明 $\mathbf{C}^4$ 上存在极小多项式等于 $z(z-1)^2$ 的算子.

对于复向量空间，这个习题给 5.21 的等价条件列表又添加了一个等价条件.

23. 设 V 是复向量空间，并且 $T \in \mathcal{L}(V)$. 证明：V 有一个由 T 的本征向量组成的基当且仅当 T 的极小多项式没有重根.

24. 设 V 是内积空间. 证明：若 $T \in \mathcal{L}(V)$ 是正规的，则 T 的极小多项式没有重根.

25. 设 $T \in \mathcal{L}(V)$，并且 $v \in V$. 设 p 为使得

$$p(T)\boldsymbol{v} = \mathbf{0}.$$

的次数最小的首一多项式. 证明 p 整除 T 的极小多项式.

26. 举例说明 $\mathbf{C}^4$ 上存在特征多项式和极小多项式都等于 $z(z-1)^2(z-3)$ 的算子.

27. 举例说明 $\mathbf{C}^4$ 上存在特征多项式等于 $z(z-1)^2(z-3)$ 而极小多项式等于 $z(z-1)(z-3)$ 的算子.

这个习题表明，每个首一多项式都是某个算子的特征多项式.

28. 设 $a_0, \cdots, a_{n-1} \in \mathbf{C}$. 求 $\mathbf{C}^n$ 上 (关于标准基的) 矩阵为

$$\begin{bmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & -a_2 \\ & & \ddots & & \vdots \\ & & & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{bmatrix}$$

的算子的极小多项式和特征多项式.

29. 设 $N \in \mathcal{L}(V)$ 是幂零的. 证明 N 的极小多项式为 z^{m+1}, 其中 m 是 N 关于任意约当基的矩阵中紧位于对角线上方的直线上 1 连续出现的最大个数.

30. 设 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$. 证明: V 不能分解成 T 的两个不变真子空间的直和当且仅当 T 的极小多项式形如 $(z-\lambda)^{\dim V}$, $\lambda \in \mathbf{C}$.

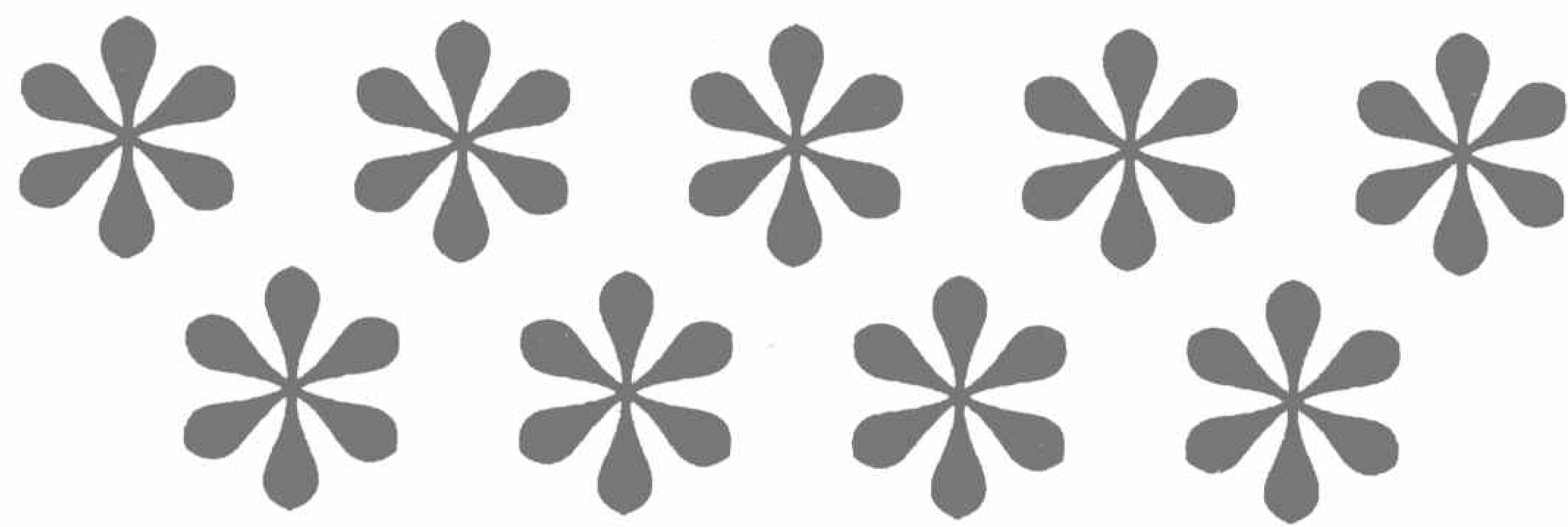
31. 设 $T \in \mathcal{L}(V)$, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基并且是 T 的约当基. 试描述 T 关于将这些 $\boldsymbol{v}$ 的顺序反过来所得到的基 $(\boldsymbol{v}_n, \cdots, \boldsymbol{v}_1)$ 的矩阵.

第 9 章　实向量空间上的算子

在这一章, 我们将深入研究实向量空间上算子的结构. 这里的结果比上一章关于复向量空间的类似结果要复杂一些.

$\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$.
V 是 $\mathbf{F}$ 上的有限维非零的向量空间.

本章有些新结果在复向量空间上也成立, 因而我们并未仅仅假设 V 是实向量空间.



§9.1 方阵的本征值

我们已经定义了算子的本征值, 现在需要把这个概念扩展到方阵. 设 $\boldsymbol{A}$ 是一个 $n \times n$ 矩阵, 它的元素都含于 $\mathbf{F}$. 一个数 $\lambda \in \mathbf{F}$ 称为 $\boldsymbol{A}$ 的 **本征值** (**eigenvalue**), 如果有非零的 $n \times 1$ 矩阵 $\boldsymbol{x}$ 使得

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}.$$

例如, 3 是 $\left[\begin{smallmatrix} 7 & 8 \\ 1 & 5 \end{smallmatrix}\right]$ 的本征值, 这是因为

$$\begin{bmatrix} 7 & 8 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 6 \\ -3 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

又如, 你应该验证, 矩阵 $\left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ 在 $\mathbf{F}$ 是 $\mathbf{R}$ 时, 没有本征值 (由定义, 本征值必须含于 $\mathbf{F}$), 而在 $\mathbf{F}$ 是 $\mathbf{C}$ 时, 有本征值 i 和 $-i$.

我们现在有两个本征值的概念 —— 一个是算子的, 另一个是方阵的. 你或许会认为这两个概念有密切联系, 这也正是我们现在要证明的.

9.1 命题: *设 $T \in \mathcal{L}(V)$, $\boldsymbol{A}$ 是 T 关于 V 的某个基的矩阵, 则 T 和 $\boldsymbol{A}$ 的本征值相同.*

证明: 设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基使得 T 关于此基有矩阵 $\boldsymbol{A}$, 并设 $\lambda \in \mathbf{F}$. 只需证明 λ 是 T 的本征值当且仅当 λ 是 $\boldsymbol{A}$ 的本征值.

首先假设 λ 是 T 的本征值, 并设非零向量 $\boldsymbol{v} \in V$ 使得 $T\boldsymbol{v} = \lambda \boldsymbol{v}$. 把 $\boldsymbol{v}$ 写成

9.2 $$\boldsymbol{v} = a_1 \boldsymbol{v}_1 + \cdots + a_n \boldsymbol{v}_n,$$

其中 $a_1, \cdots, a_n \in \mathbf{F}$. 设 $\boldsymbol{x}$ 是向量 $\boldsymbol{v}$ 关于基 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 的矩阵. 回想一下, 在第 3 章中这指的是

9.3 $$\boldsymbol{x} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

因此,

$$\boldsymbol{A}\boldsymbol{x} = \mathcal{M}(T)\mathcal{M}(\boldsymbol{v}) = \mathcal{M}(T\boldsymbol{v}) = \mathcal{M}(\lambda\boldsymbol{v}) = \lambda\mathcal{M}(\boldsymbol{v}) = \lambda\boldsymbol{x},$$

其中第二个等式是由 3.14 得到的. 上式表明, λ 是 $\boldsymbol{A}$ 的本征值.

为了证明另一个方向的蕴含关系, 现在假设 λ 是 $\boldsymbol{A}$ 的本征值, 并设 $\boldsymbol{x}$ 是非零的 $n \times 1$ 矩阵, 使得 $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. 把 $\boldsymbol{x}$ 写成 9.3 的形式, 其中标量 $a_1, \cdots, a_n \in \mathbf{F}$. 定义 $\boldsymbol{v} \in V$ 如 9.2, 则

$$\mathcal{M}(T\boldsymbol{v}) = \mathcal{M}(T)\mathcal{M}(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} = \mathcal{M}(\lambda\boldsymbol{v}),$$

其中第一个等式成立是由于 3.14. 由上式得 $T\boldsymbol{v} = \lambda\boldsymbol{v}$, 故 λ 是 T 的本征值. ■

因为每个方阵都是某个算子的矩阵, 所以利用上面这个命题可以把算子本征值的结果翻译成方阵本征值的语言. 例如, 每个复方阵都有本征值 (由 5.10). 又如, 每个 $n \times n$ 矩阵最多有 n 个互不相同的本征值 (由 5.9).

§9.2 分块上三角矩阵

前面我们证明了, 复向量空间上每个算子关于某个基都有上三角矩阵 (参见 5.13), 在这一节我们会看到在实向量空间上也几乎如此.

在前面两章中, 我们用到了分块对角矩阵, 它是对角矩阵的推广. 现在我们需要上三角矩阵的相应推广. **分块上三角矩阵 (block upper triangular matrix)** 是如下形式的方阵

我们照例用星号 $*$ 表示对所考虑的问题不太起作用的矩阵元素.

$$\begin{bmatrix} \boldsymbol{A}_1 & & * \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 都是方阵, 它们位于对角线上, 而 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 下方的元素全等于 0, 并且 $*$ 表示任意元素. 例如, 矩阵

$$A = \begin{bmatrix} 4 & 10 & 11 & 12 & 13 \\ 0 & -3 & -3 & 14 & 25 \\ 0 & -3 & -3 & 16 & 17 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 5 & 5 \end{bmatrix}$$

是分块上三角矩阵:

$$A = \begin{bmatrix} A_1 & & * \\ & A_2 & \\ 0 & & A_3 \end{bmatrix},$$

其中

$$A_1 = \begin{bmatrix} 4 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -3 & -3 \\ -3 & -3 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}.$$

我们现在证明, 对于实向量空间上的每个算子, 都有基使得该算子关于此基具有分块上三角矩阵, 并且对角线上块的大小不超过 2×2.

每个上三角矩阵都是分块上三角矩阵, 其中对角线上都是 1×1 的块. 另一个极端的情形是, 每个方阵都是分块上三角矩阵, 这是因为我们可以取第一个(也是唯一的)块是整个矩阵. 块越小越好, 因为块越小矩阵中的 0 就越多.

9.4 定理: 设 V 是实向量空间, 并设 $T \in \mathcal{L}(V)$, 则 V 有基使得 T 关于此基有分块上三角矩阵

9.5
$$\begin{bmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$

其中每个 A_j 都是 1×1 矩阵或没有本征值的 2×2 矩阵.

证明: 若 $\dim V = 1$, 则结果显然成立.

下面考虑 $\dim V = 2$ 的情形. 若 T 有本征值 λ, 则取 $v_1 \in V$ 为任意非零本征向量. 把 (v_1) 扩充成 V 的基 (v_1, v_2). T 关于此基有如下形式的上三角矩阵

$$\begin{bmatrix} \lambda & a \\ 0 & b \end{bmatrix}.$$

特别地, 若 T 有本征值, 则 V 有基使得 T 关于此基有上三角矩阵. 若 T 没有本征值, 则任取 V 的一个基 (v_1, v_2). 关于此基, T

的矩阵没有本征值 (由 9.1). 于是, 在 $\dim V = 2$ 时, 无论 T 是否有本征值, 结论总成立.

现在假设 $\dim V > 2$, 并设结果对于维数更小的实向量空间都成立. 如果 T 有本征值, 则取 U 是 V 的在 T 下不变的 1 维子空间; 否则取 U 是 V 的在 T 下不变的 2 维子空间 (5.24 确保了总可以这样取). 任取 U 的一个基, 并设 $\boldsymbol{A}_1$ 是 $T|_U$ 关于此基的矩阵. 若 $\boldsymbol{A}_1$ 是 2×2 矩阵, 则 T 没有本征值 (否则 U 就取成 1 维的了), 故 $T|_U$ 没有本征值. 因此, 若 $\boldsymbol{A}_1$ 是 2×2 矩阵, 则 $\boldsymbol{A}_1$ 没有本征值 (参见 9.1).

设 W 是 V 的任意子空间, 使得

$$V = U \oplus W;$$

2.13 确保了这样的 W 一定存在. 因为 W 的维数比 V 小, 所以我们想对 $T|_W$ 用归纳法假设. 然而, W 未必在 T 下不变, 即 $T|_W$ 未必是 W 上的算子. 将 T 与投影 $P_{W,U}$ 复合即可得 W 上的算子. 具体地, 定义 $S \in \mathcal{L}(W)$ 如下

回想一下, 若 $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{u}$, 其中 $\boldsymbol{w} \in W$, $\boldsymbol{u} \in U$, 则 $P_{W,U}\boldsymbol{v} = \boldsymbol{w}$.

$$S\boldsymbol{w} = P_{W,U}(T\boldsymbol{w}), \quad \boldsymbol{w} \in W.$$

注意到对每个 $\boldsymbol{w} \in W$ 都有

9.6
$$\begin{aligned} T\boldsymbol{w} &= P_{U,W}(T\boldsymbol{w}) + P_{W,U}(T\boldsymbol{w}) \\ &= P_{U,W}(T\boldsymbol{w}) + S\boldsymbol{w}. \end{aligned}$$

由归纳法假设, W 有基使得 S 关于此基有如下形式的分块上三角矩阵

$$\begin{bmatrix} \boldsymbol{A}_2 & & * \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中每个 $\boldsymbol{A}_j$ 都是 1×1 矩阵或没有本征值的 2×2 矩阵. 把 W 的这个基添加到上面取定的 U 的基, 可得 V 的基. 稍加思索即知 (用 9.6), T 关于此基的矩阵是形如 9.5 的分块上三角矩阵.

■

§9.3　特征多项式

对于复向量空间上的算子, 我们定义过特征多项式, 并利用上三角矩阵阐述了它的性质. 本节要对实向量空间上的算子做同样的事情, 只不过必须使用上一个定理提供的分块上三角矩阵而不是使用上三角矩阵.

在上一章, 我们并未对复方阵定义特征多项式, 这是因为我们强调的是算子而非矩阵. 然而, 为了理解实向量空间上的算子, 我们需要定义 1×1 和 2×2 实矩阵的特征多项式. 再利用对角线上块大小不超过 2×2 的分块上三角矩阵, 我们就可以定义实向量空间上算子的特征多项式.

为了给方阵的特征多项式的定义提供动机, 我们希望下面的陈述是真的 (回想凯莱–哈密顿定理; 参见 8.20): 如果 $T \in \mathcal{L}(V)$关于 V 的某个基有矩阵 $\boldsymbol{A}$, 并且 q 是 $\boldsymbol{A}$ 的特征多项式, 则 $q(T) = 0$.

先考虑 1×1 矩阵这种平凡情形. 设 V 是 1 维实向量空间, 并且 $T \in \mathcal{L}(V)$. 若 $[\lambda]$ 等于 T 关于 V 的某个基的矩阵, 则 T 等于 λI. 于是, 如果令 q 是 1 次多项式 $q(x) = x - \lambda$, 则 $q(T) = 0$. 因此, 可以定义 $[\lambda]$ 的特征多项式为 $x - \lambda$.

现在考虑 2×2 实矩阵. 设 V 是 2 维实向量空间, $T \in \mathcal{L}(V)$, 并且

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

是 T 关于 V 的某个基 $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ 的矩阵. 我们来求一个 2 次的首一多项式 q 使得 $q(T) = 0$. 若 $b = 0$, 则上述矩阵是上三角的. 假如我们处理的是复向量空间, 那么我们知道 T 有特征多项式 $(z - a)(z - d)$. 于是, $(x - a)(x - d)$ 是一个合理的候选, 其中用 x 代替 z 是为了强调现在考虑的是实向量空间. 来看看当 $b \neq 0$

时, 多项式 $(x-a)(x-d)$ 作用到 T 是否得 0. 我们有

$$\begin{aligned}(T-aI)(T-dI)\boldsymbol{v}_1&=(T-dI)(T-aI)\boldsymbol{v}_1\\&=(T-dI)(b\boldsymbol{v}_2)=bc\boldsymbol{v}_1\end{aligned}$$

以及

$$(T-aI)(T-dI)\boldsymbol{v}_2=(T-aI)(c\boldsymbol{v}_1)-bc\boldsymbol{v}_2.$$

因此, 只要 $bc\neq 0$, 则 $(T-aI)(T-dI)$ 一定不等于 0. 然而, 上式表明, $(T-aI)(T-dI)-bcI=0$(因为这个算子在基上等于 0, 故在 V 上必为 0). 于是, 若 $q(x)=(x-a)(x-d)-bc$, 则 $q(T)=0$.

受前一段的启发, 我们把 2×2 矩阵 $\left[\begin{smallmatrix}a&c\\b&d\end{smallmatrix}\right]$ 的**特征多项式 (characteristic polynomial)** 定义为 $(x-a)(x-d)-bc$. 此处我们只讨论实矩阵. 下一个结果表明, 我们已经找到了 2×2 矩阵的特征多项式的唯一合理定义.

9.7 命题: *设 V 是 2 维实向量空间, $T\in\mathcal{L}(V)$ 没有本征值, $p\in\mathcal{P}(\mathbf{R})$ 是 2 次首一多项式, $\boldsymbol{A}$ 是 T 关于 V 的某个基的矩阵.*

(a) 若 p 等于 $\boldsymbol{A}$ 的特征多项式, 则 $p(T)=0$.

(b) 若 p 不等于 $\boldsymbol{A}$ 的特征多项式, 则 $p(T)$ 可逆.

如果去掉 T 没有本征值的假设, 这个命题中的 (b) 就不成立. 例如, 对于由 $T(x_1,x_2)=(0,x_2)$ 定义的 $T\in\mathcal{L}(\mathbf{R}^2)$, 取 $p(x)=x(x-2)$, 则 p 不是 T 关于标准基的矩阵的特征多项式, 但 $p(T)$ 不可逆.

证明: 在上面的讨论中, 我们已经证明了 (a). 为证 (b), 设 q 表示 $\boldsymbol{A}$ 的特征多项式, 并设 $p\neq q$. 令 $p(x)=x^2+\alpha_1x+\beta_1$, $q(x)=x^2+\alpha_2x+\beta_2$, 其中 $\alpha_1,\beta_1,\alpha_2,\beta_2\in\mathbf{R}$. 现在

$$p(T)=p(T)-q(T)=(\alpha_1-\alpha_2)T+(\beta_1-\beta_2)I.$$

若 $\alpha_1=\alpha_2$, 则 $\beta_1\neq\beta_2$ (否则, $p=q$), 于是 $p(T)$ 是恒等映射的非零标量倍, 因此是可逆的. 若 $\alpha_1\neq\alpha_2$, 则由 T 没有本征值知

$$p(T)=(\alpha_1-\alpha_2)\left(T-\frac{\beta_2-\beta_1}{\alpha_1-\alpha_2}I\right),$$

是可逆算子. 于是, (b) 成立. ■

设 V 是 2 维实向量空间, 并设 $T \in \mathcal{L}(V)$ 没有本征值. 上一个命题证明了, 恰有一个 2 次首一多项式, 把它作用到 T 得 0. 于是, 虽然 T 关于不同的基可能有不同的矩阵, 但这些矩阵都有相同的特征多项式. 例如, 考虑如下定义的 $T \in \mathcal{L}(\mathbf{R}^2)$

9.8 $$T(x_1, x_2) = (3x_1 + 5x_2, -2x_1 - x_2).$$

T 关于 $\mathbf{R}^2$ 的标准基的矩阵是

$$\begin{bmatrix} 3 & 5 \\ -2 & -1 \end{bmatrix}.$$

这个矩阵的特征多项式是 $(x-3)(x+1)+2\times 5$, 即 x^2-2x+7. 你应该验证, T 关于基 $((-2,1),(1,2))$ 的矩阵等于

$$\begin{bmatrix} 1 & -6 \\ 1 & 1 \end{bmatrix}.$$

这个矩阵的特征多项式是 $(x-1)(x-1)+1\times 6$, 也等于 x^2-2x+7, 这与使用标准基所得到的结果相同.

在分析复向量空间 V 上的算子 T 的上三角矩阵时, 我们发现如下形式的子空间

$$\operatorname{null}(T-\lambda I)^{\dim V}$$

发挥了关键作用 (参见 8.10). 这些子空间在研究实向量空间上的算子时也会发挥作用, 但是由于我们必须考虑带 2×2 块的分块上三角矩阵, 从而如下形式的子空间

$$\operatorname{null}(T^2+\alpha T+\beta I)^{\dim V}$$

也将发挥关键作用. 先来考虑 1 维和 2 维实向量空间.

首先设 V 是 1 维实向量空间, 并设 $T \in \mathcal{L}(V)$. 若 $\lambda \in \mathbf{R}$, 则 $\operatorname{null}\mathcal{L}(T-\lambda I)$ 在 λ 是 T 的本征值时等于 V, 否则等于 $\{\mathbf{0}\}$. 如果 $\alpha, \beta \in \mathbf{R}$ 使得 $\alpha^2 < 4\beta$, 那么

$$\operatorname{null}(T^2+\alpha T+\beta I)=\{\mathbf{0}\}.$$

(证明: 因为 V 是 1 维的, 故有常数 $\lambda\in\mathbf{R}$ 使得对所有 $\boldsymbol{v}\in V$ 都有 $T\boldsymbol{v}=\lambda\boldsymbol{v}$. 于是, $(T^2+\alpha T+\beta I)\boldsymbol{v}=(\lambda^2+\alpha\lambda+\beta)\boldsymbol{v}$, 而由不等式 $\alpha^2<4\beta$ 得 $\lambda^2+\alpha\lambda+\beta\neq 0$, 故 $\operatorname{null}\mathcal{L}(T^2+\alpha T+\beta I)=\{\mathbf{0}\}$.)

回想一下, $\alpha^2<4\beta$ 意味着 $x^2+\alpha x+\beta$ 没有实根; 参见 4.11.

现在设 V 是 2 维实向量空间, 并设 $T\in\mathcal{L}(V)$ 没有本征值. 若 $\lambda\in\mathbf{R}$, 则 $\operatorname{null}\mathcal{L}(T-\lambda I)$ 等于 $\{\mathbf{0}\}$(因为 T 没有本征值). 若 $\alpha,\beta\in\mathbf{R}$ 使得 $\alpha^2<4\beta$, 则当 $x^2+\alpha x+\beta$ 是 T 关于 V 的某个 (或等价地, 每个) 基的矩阵的特征多项式时, $\operatorname{null}\mathcal{L}(T^2+\alpha T+\beta I)$ 等于 V, 否则等于 $\{\mathbf{0}\}$ (由 9.7). 注意, 对于这个算子, 只有两种极端的情况 —— $T^2+\alpha T+\beta I$ 的零空间要么是 $\{\mathbf{0}\}$, 要么是整个空间; 它不可能是 1 维的.

现在设 V 是任意维数的实向量空间, 并设 $T\in\mathcal{L}(V)$. 我们知道 V 有基使得 T 关于此基有分块上三角矩阵, 其中对角线上块的大小都不超过 2×2 (参见 9.4). 一般地, 这个矩阵不唯一 —— V 有很多不同的基使得 T 有这样的分块上三角矩阵, 并且关于这些不同的基我们可以得到不同的分块上三角矩阵.

我们在处理复向量空间和上三角矩阵时遇到过类似的情况. 那时虽然关于不同的基得到了不同的上三角矩阵, 但这些矩阵对角线上的元素却总是相同的 (尽管次序可能不同). 类似的性质对于实向量空间和分块上三角矩阵是否也成立呢? 具体地, 一个给定的 2×2 矩阵在 T 的分块上三角矩阵的对角线上出现的次数是否与基的选取无关呢? 很遗憾, 这个问题的答案是否定的. 例如, 我们以前看到过, 由 9.8 定义的算子 $T\in\mathcal{L}(\mathbf{R}^2)$ 就有两个不同的 2×2 矩阵.

虽然一个给定的 2×2 矩阵在 T 的分块上三角矩阵的对角线上出现的次数可能与所取的基有关, 但是, 如果考虑的是特征多项式而不是实际的矩阵, 我们就会发现一个给定的特征多项式出现的次数与基的选取无关. 这就是下面定理的内容, 它将成为分析实向量空间上算子结构的关键工具.

9.9 定理：设 V 是实向量空间，$T \in \mathcal{L}(V)$，并设 T 关于 V 的某个基的矩阵是

9.10
$$\begin{bmatrix} \boldsymbol{A}_1 & & * \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix}.$$

其中每个 $\boldsymbol{A}_j$ 都是 1×1 矩阵或没有本征值的 2×2 矩阵.

(a) 若 $\lambda \in \mathbf{R}$，则 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 中恰好有 $\dim \operatorname{null} \mathcal{L}(T - \lambda I)^{\dim V}$ 个矩阵等于 1×1 矩阵 $[\lambda]$.

(b) 若 $\alpha, \beta \in \mathbf{R}$ 满足 $\alpha^2 < 4\beta$，则在 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 中恰好有

$$\frac{\dim \operatorname{null}(T^2 + \alpha T + \beta I)^{\dim V}}{2}$$

个矩阵的特征多项式等于 $x^2 + \alpha x + \beta$.

这个结果表明 $\operatorname{null} \mathcal{L}(T^2 + \alpha T + \beta I)^{\dim V}$ 的维数一定是偶数.

这个证明与复向量空间类似结果 (8.10) 的证明所用的思想相同. 实的情形通常更复杂一点，但无需创新.

证明：我们将构造一个对 (a) 和 (b) 都适用的证明. 为此，设 $\lambda, \alpha, \beta \in \mathbf{R}$ 使得 $\alpha^2 < 4\beta$. 定义 $p \in \mathcal{P}(\mathbf{R})$ 如下

$$p(x) = \begin{cases} x - \lambda & \text{若要证明 (a);} \\ x^2 + \alpha x + \beta & \text{若要证明 (b).} \end{cases}$$

设 d 表示 p 的次数. 若要证明 (a)，则 $d = 1$；若要证明 (b)，则 $d = 2$.

对矩阵 9.10 对角线上的块数 m 用归纳法. 若 $m = 1$，则 $\dim V = 1$ 或 $\dim V = 2$；定理前面的讨论说明这时要证的结果成立. 于是，假设 $m > 1$，并设要证的结果对 $m - 1$ 成立.

为了方便，设 $n = \dim V$. 考虑 V 的一个基使得 T 关于此基有分块上三角矩阵 9.10. 令 U_j 表示相应于 $\boldsymbol{A}_j$ 的那些基向量张成的子空间. 若 $\boldsymbol{A}_j$ 是 1×1 矩阵，则 $\dim U_j = 1$；若 $\boldsymbol{A}_j$ 是 2×2 矩阵，则 $\dim U_j = 2$. 令 $U = U_1 + \cdots + U_{m-1}$. 显然，$U$ 在 T 下是不变的，并且 $T|_U$ 关于那个明显的基（由相应于 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{m-1}$ 的基向量组成）的矩阵是

9.11
$$\begin{bmatrix} \boldsymbol{A}_1 & & * \\ & \ddots & \\ 0 & & \boldsymbol{A}_{m-1} \end{bmatrix}.$$

于是, 由归纳法假设,

9.12 $A_1, \cdots, A_{m-1}$ 中特征多项式等于 p 的矩阵恰好有 $(1/d)\dim\operatorname{null} p(T|_U)^n$ 个.

实际上, 在 9.12 中由归纳法假设所得的指数是 $\dim U$ 而不是 n, 但是, 可以用 n 代替 $\dim U$ (由 8.6), 从而可得上面的陈述.

设 $\boldsymbol{u}_m \in U_m$, 并设算子 $S \in \mathcal{L}(U_m)$ (关于 U_m 的相应基) 的矩阵为 $\boldsymbol{A}_m$. 特别地, $S\boldsymbol{u}_m = P_{U_m,U}T\boldsymbol{u}_m$. 现在

$$\begin{aligned} T\boldsymbol{u}_m &= P_{U,U_m}T\boldsymbol{u}_m + P_{U_m,U}T\boldsymbol{u}_m \\ &= *_U + S\boldsymbol{u}_m, \end{aligned}$$

其中 $*_U$ 表示 U 中的一个向量. 注意到 $S\boldsymbol{u}_m \in U_m$; 故把 T 作用到上式的两端可得

$$T^2\boldsymbol{u}_m = *_U + S^2\boldsymbol{u}_m,$$

其中 $*_U$ 仍表示 U 中的一个向量, 但可能是和前一个 $*_U$ 不同的向量 (记号 $*_U$ 用以强调这是 U 中的一个向量, 但我们并不关心它是哪一个特殊向量 —— 每次使用记号 $*_U$ 时, 它都可以表示 U 中的不同向量). 上面这两个等式证明了, 对某个 $*_U \in U$ 有

9.13 $$p(T)\boldsymbol{u}_m = *_U + p(S)\boldsymbol{u}_m.$$

注意到 $p(S)\boldsymbol{u}_m \in U_m$; 于是, 反复使用上式进行迭代可知, 对某个 $*_U \in U$ 有

9.14 $$p(T)^n\boldsymbol{u}_m = *_U + p(S)^n\boldsymbol{u}_m.$$

现在我们分两种情况来证明. 首先考虑 $\boldsymbol{A}_m$ 的特征多项式不等于 p 的情况. 我们将证明此时有

9.15 $$\operatorname{null} p(T)^n \subset U.$$

一旦如此, 即知

$$\operatorname{null} p(T)^n = \operatorname{null} p(T|_U)^n,$$

故由 9.12 可知, 在 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 中恰好有 $(1/d)\dim\operatorname{null} p(T)^n$ 个矩阵的特征多项式为 p, 这就完成了 $\boldsymbol{A}_m$ 的特征多项式不等于 p 这种情形的证明.

为证 9.15 (仍设 $\boldsymbol{A}_m$ 的特征多项式不等于 p), 设 $\boldsymbol{v} \in \operatorname{null} p(T)^n$, 并将 $\boldsymbol{v}$ 写成 $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{u}_m$, 其中 $\boldsymbol{u} \in U$, $\boldsymbol{u}_m \in U_m$. 由 9.14 可知, 对某个 $*_U \in U$ 有

$$\mathbf{0} = p(T)^n \boldsymbol{v} = p(T)^n \boldsymbol{u} + p(T)^n \boldsymbol{u}_m = p(T)^n \boldsymbol{u} + *_U + p(S)^n \boldsymbol{u}_m.$$

因为向量 $p(T)^n \boldsymbol{u}$ 和 $*_U$ 都在 U 中, 而 $p(S)^n \boldsymbol{u}_m \in U_m$, 所以 $p(S)^n \boldsymbol{u}_m = \mathbf{0}$. 然而, $p(S)$ 可逆 (参见此定理前面关于 1 维和 2 维子空间的讨论, 并注意到 $\dim U_m \leqslant 2$), 故 $\boldsymbol{u}_m = \mathbf{0}$. 于是 $\boldsymbol{v} = \boldsymbol{u} \in U$, 这就证明了 9.15.

现在考虑 $\boldsymbol{A}_m$ 的特征多项式等于 p 的情形. 注意到此时有 $\dim U_m = d$. 根据 9.12, 只需证明

9.16
$$\dim \operatorname{null} p(T)^n = \dim \operatorname{null} p(T|_U)^n + d.$$

利用两个子空间之和的维数公式 (2.18) 可得

$$\begin{aligned}\dim \operatorname{null} p(T)^n &= \dim(U \cap \operatorname{null} p(T)^n) \\ &\quad + \dim(U + \operatorname{null} p(T)^n) - \dim U \\ &= \dim \operatorname{null} p(T|_U)^n \\ &\quad + \dim(U + \operatorname{null} p(T)^n) - (n-d).\end{aligned}$$

若 $U + \operatorname{null} p(T)^n = V$, 则 $\dim(U + \operatorname{null} p(T)^n) = n$, 将其与上面关于 $\dim \operatorname{null} p(T)^n$ 的公式联合起来即可得到 9.16. 于是, 只需证明 $U + \operatorname{null} p(T)^n = V$.

为证 $U + \operatorname{null} p(T)^n = V$, 设 $\boldsymbol{u}_m \in U_m$. 因为 S 的矩阵 (即 $\boldsymbol{A}_m$) 有特征多项式 p, 所以 $p(S) = 0$. 于是, $p(T)\boldsymbol{u}_m \in U$ (由 9.13). 现在

$$\begin{aligned}p(T)^n \boldsymbol{u}_m &= p(T)^{n-1}(p(T)\boldsymbol{u}_m) \in \operatorname{range} p(T|_U)^{n-1} \\ &= \operatorname{range} p(T|_U)^n,\end{aligned}$$

其中最后一个等式是根据 8.9 得到的. 因此, 可取 $\boldsymbol{u} \in U$ 使得 $p(T)^n \boldsymbol{u}_m = p(T|_U)^n \boldsymbol{u}$. 现在

$$\begin{aligned}p(T)^n(\boldsymbol{u}_m - \boldsymbol{u}) &= p(T)^n \boldsymbol{u}_m - p(T)^n \boldsymbol{u} \\ &= p(T)^n \boldsymbol{u}_m - p(T|_U)^n \boldsymbol{u} \\ &= \mathbf{0}.\end{aligned}$$

于是, $\boldsymbol{u}_m - \boldsymbol{u} \in \operatorname{null} p(T)^n$, 故 $\boldsymbol{u}_m$ 等于 $\boldsymbol{u} + (\boldsymbol{u}_m - \boldsymbol{u})$, 包含于 $U + \operatorname{null} p(T)^n$. 也就是说, $U_m \subset U + \operatorname{null} p(T)^n$. 因此, $V = U + U_m \subset U + \operatorname{null} p(T)^n$, 故 $U + \operatorname{null} p(T)^n = V$. ■

我们在上一章曾看到, 复向量空间上算子的本征值为分析算子结构提供了主要工具. 在实向量空间上, 一个算子本征值(计算重数) 的个数可能少于向量空间维数. 前一个定理提示我们可以定义一个概念来弥补这个不足. 我们将看到下一段所给出的定义有助于使得实向量空间上的算子理论与复向量空间上的算子理论更相似.

设 V 是实向量空间, 并设 $T \in \mathcal{L}(V)$. 如果 $\alpha^2 < 4\beta$, 并且

$$T^2 + \alpha T + \beta I$$

不是单的, 则称有序实数对 (α, β) 为 T 的**本征对 (eigenpair)**. 前一个定理证明了, T 只能有有限多个本征对, 这是因为每个本征对都对应 9.10 对角线上一个 2×2 矩阵的特征多项式, 而在对角线上只有有限个这样的矩阵. 按照 9.9, 我们把 T 的本征对 (α, β) 的**重数 (multiplicity)** 定义为

尽管选用了**本征对**这个词以便和**本征值**这个词保持一致, 但这个术语并不流行.

$$\frac{\dim \operatorname{null}(T^2 + \alpha T + \beta I)^{\dim V}}{2}.$$

由 9.9 可知, (α, β) 的重数等于 9.10 的对角线上以 $x^2 + \alpha x + \beta$ 为特征多项式的 2×2 矩阵的个数.

例如, 考虑算子 $T \in \mathcal{L}(\mathbf{R}^3)$, 它 (关于标准基) 的矩阵等于

$$\begin{bmatrix} 3 & -1 & -2 \\ 3 & 2 & -3 \\ 1 & 2 & 0 \end{bmatrix}$$

你应该验证, $(-4, 13)$ 是 T 的 1 重本征对; 注意到 $T^2 - 4T + 13I$ 不是单的, 这是因为 $(-1, 0, 1)$ 和 $(1, 1, 0)$ 都在它的零空间中. 即使不做计算, 你也应该能验证, T 没有其他本征对 (用 9.9). 你还应该验证, 1 是 T 的 1 重本征值, 有本征向量 $(1, 0, 1)$, 并且 T 没有其他本征值.

在上面的例子中, T 的本征值重数之和加上 T 的本征对重数的 2 倍等于 3, 这是 T 的定义域的维数. 下一个命题指出, 这在实向量空间上总成立.

这个命题表明, 虽然实向量空间上的算子可以没有本征值或没有本征对, 但不能同时没有这两个有用的对象. 它还表明, 实向量空间 V 上的算子最多有 $(\dim V)/2$ 个互不相同的本征对.

9.17 命题: 若 V 是实向量空间, 并且 $T \in \mathcal{L}(V)$, 则 T 的所有本征值重数之和加上 T 的所有本征对重数的 2 倍之和等于 $\dim V$.

证明: 设 V 是实向量空间, 并设 $T \in \mathcal{L}(V)$, 则 V 有基使得 T 关于此基的矩阵如 9.10. 本征值 λ 的重数等于 1×1 矩阵 $[\lambda]$ 在这个矩阵对角线上出现的次数 (根据 9.9). 本征对 (α, β) 的重数等于这个矩阵对角线上以 $x^2 + \alpha x + \beta$ 为特征多项式的 2×2 矩阵的个数 (根据 9.9). 因为这个矩阵的对角线的长度等于 $\dim V$, 所以 T 的所有本征值的重数之和再加上 T 的所有本征对的重数的 2 倍之和一定等于 $\dim V$. ■

设 V 是实向量空间, 并设 $T \in \mathcal{L}(V)$. T 关于 V 的某个基有如下形式的分块上三角矩阵

9.18
$$\begin{bmatrix} \boldsymbol{A}_1 & & * \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中每个 $\boldsymbol{A}_j$ 都是 1×1 矩阵或没有本征值的 2×2 矩阵 (参见 9.4). 定义 T 的**特征多项式 (characteristic polynomial)** 为 $\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m$ 的特征多项式之积. 明确地说, 对于每个 j, 定义 $q_j \in \mathcal{P}(\mathbf{R})$ 如下

9.19
$$q_j(x) = \begin{cases} x - \lambda & \text{若 } \boldsymbol{A}_j \text{ 等于 } [\lambda]; \\ (x-a)(x-d) - bc & \text{若 } \boldsymbol{A}_j \text{ 等于 } \begin{bmatrix} a & c \\ b & d \end{bmatrix}, \end{cases}$$

注意到和在复向量空间上一样, T 的特征多项式的根等于 T 的本征值.

则 T 的特征多项式是

$$q_1(x) \cdots q_m(x).$$

显然, T 的特征多项式的次数为 $\dim V$. 进一步, 9.9 确保 T 的特征多项式只与 T 有关, 而与特殊基的选取无关. 现在我们能

证明上一章许诺过的一个结果, 上一章我们对复向量空间上的算子证明过类似的定理 (8.20).

9.20 凯莱 – 哈密顿定理 (Cayley-Hamilton Theorem): 设 V 是实向量空间, $T \in \mathcal{L}(V)$, q 表示 T 的特征多项式, 则 $q(T) = 0$.

证明: 取 V 的一个基使得 T 关于此基有形如 9.18 的分块上三角矩阵, 其中每个 $\boldsymbol{A}_j$ 都是 1×1 矩阵或没有本征值的 2×2 矩阵. 设 U_j 是相应于 $\boldsymbol{A}_j$ 的基向量张成的 1 维或 2 维子空间. 定义 q_j 如 9.19. 为证 $q(T) = 0$, 只需证明 $q(T)|_{U_j} = 0$, $j = 1, \cdots, m$. 为此, 只需证明

这个证明与复向量空间上类似结果 (8.20) 的证明所用的思想相同.

9.21 $$q_1(T)\cdots q_j(T)|_{U_j} = 0, \quad j = 1, \cdots, m.$$

对 j 用归纳法来证明 9.21. 先设 $j = 1$. 因为 $\mathcal{M}(T)$ 是 9.18 给出的矩阵, 所以 $q_1(T)|_{U_1} = 0$ (若 $\dim U_1 = 1$, 则显然; 若 $\dim U_1 = 2$, 则根据 9.7(a)), 从而, 当 $j = 1$ 时, 9.21 成立.

现在假设 $1 < j \leqslant m$, 并设

$$0 = q_1(T)|_{U_1}$$

$$0 = q_1(T)q_2(T)|_{U_2}$$

$$\vdots$$

$$0 = q_1(T)\cdots q_{j-1}(T)|_{U_{j-1}}.$$

若 $\boldsymbol{v} \in U_j$, 则根据 9.18 可得

$$q_j(T)\boldsymbol{v} = \boldsymbol{u} + q_j(S)\boldsymbol{v},$$

其中 $\boldsymbol{u} \in U_1 + \cdots + U_{j-1}$, 并且 $S \in \mathcal{L}(U_j)$ 有特征多项式 q_j. 因为 $q_j(S) = 0$ (若 $\dim U_j = 1$, 则显然; 若 $\dim U_j = 2$, 则根据 9.7(a)), 所以上式表明, 当 $\boldsymbol{v} \in U_j$ 时, 必有

$$q_j(T)\boldsymbol{v} \in U_1 + \cdots + U_{j-1}.$$

于是, 对于 $\boldsymbol{v} \in U_j$, 由归纳法假设, 把 $q_1(T)\cdots q_{j-1}(T)$ 作用到 $q_j(T)\boldsymbol{v}$ 得 $\boldsymbol{0}$. 即 9.21 成立. ■

设 V 是实向量空间, 并设 $T \in \mathcal{L}(V)$. 显然, 凯莱–哈密顿定理 (9.20) 表明, 和复向量空间上的情形一样, T 的极小多项式的次数不超过 $\dim V$. 如果 T 的极小多项式的次数等于 $\dim V$, 则仍和复向量空间上的情形一样, T 的极小多项式一定等于 T 的特征多项式. 这可以由凯莱–哈密顿定理 (9.20) 和 8.34 得到.

最后, 我们可以证明实向量空间上的算子的一个主要结构定理. 应该把下面这个定理同复向量空间上的类似结果 8.23 对比一下.

m 或 M 都有可能是 0.

9.22 定理: 设 V 是实向量空间, $T \in \mathcal{L}(V)$, $\lambda_1, \cdots, \lambda_m$ 是 T 的所有互不相同的本征值, $U_1, \cdots, U_m$ 分别是相应的广义本征向量的集合, $(\alpha_1, \beta_1), \cdots, (\alpha_M, \beta_M)$ 是 T 的所有互不相同的本征对, $V_j = \operatorname{null}(T^2 + \alpha_j T + \beta_j I)^{\dim V}$, 则

(a) $V = U_1 \oplus \cdots \oplus U_m \oplus V_1 \oplus \cdots \oplus V_M$;

(b) 每个 U_j 和每个 V_j 在 T 下都是不变的;

(c) 每个 $(T - \lambda_j I)|_{U_j}$ 和每个 $(T^2 + \alpha_j T + \beta_j I)|_{V_j}$ 都是幂零的.

这个证明与复向量空间上类似结果 (8.20) 的证明所用的思想相同.

证明: 根据 8.22 即可得 (b). 根据定义显然可得 (c).

为证 (a), 回想一下, $\dim U_j$ 等于 T 的本征值 λ_j 的重数, 并且 $\dim V_j$ 等于 T 的本征对 (α_j, β_j) 的重数的 2 倍. 于是, 由 9.17 可得

9.23
$$\begin{aligned}\dim V = &\dim U_1 + \cdots + \dim U_m \\ &+ \dim V_1 + \cdots + \dim V_M.\end{aligned}$$

设 $U = U_1 + \cdots + U_m + V_1 + \cdots + V_M$. 注意到 U 在 T 下是不变的. 于是, 可以定义 $S \in \mathcal{L}(U)$ 如下

$$S = T|_U.$$

因为 T 的所有广义本征向量都在 S 的定义域 U 中, 所以 S 和 T 有同样的本征值以及同样的重数. 类似地, S 和 T 也有同样的本征对以及同样的重数. 于是, 把 9.17 作用于 S 可得

$$\dim U = \dim U_1 + \cdots + \dim U_m + \dim V_1 + \cdots + \dim V_M.$$

再结合 9.23 可得 $\dim V = \dim U$. 因为 U 是 V 的子空间, 所以

$V = U$. 也就是说,

$$V = U_1 + \cdots + U_m + V_1 + \cdots + V_M.$$

利用 9.23 并结合 2.19 即可知 (a) 成立. ■

习 题

1. 证明每行元素之和都等于 1 的方阵必有一个本征值等于 1.

2. 考虑 2×2 实矩阵

$$\boldsymbol{A}=\begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

证明: $\boldsymbol{A}$(在 $\mathbf{R}$ 中) 有本征值当且仅当

$$(a-d)^2+4bc\geqslant 0.$$

3. 设 $\boldsymbol{A}$ 是分块对角矩阵

$$\boldsymbol{A}=\begin{bmatrix} \boldsymbol{A}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中每个 $\boldsymbol{A}_j$ 都是方阵. 证明 $\boldsymbol{A}$ 的本征值之集等于 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_m$ 的本征值之集的并.

显然, 习题 4 的结论要比习题 3 强. 尽管如此, 你可能还是想先做习题 3, 因为它比习题 4 简单.

4. 设 $\boldsymbol{A}$ 是分块上三角矩阵

$$\boldsymbol{A}=\begin{bmatrix} \boldsymbol{A}_1 & & * \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中每个 $\boldsymbol{A}_j$ 都是方阵. 证明 $\boldsymbol{A}$ 的本征值之集等于 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_m$ 的本征值之集的并.

5. 设 V 是实向量空间, $T\in\mathcal{L}(V)$, 并且 $\alpha,\beta\in\mathbf{R}$ 使得 $T^2+\alpha T+\beta I=0$. 证明 T 有本征值当且仅当 $\alpha^2\geqslant 4\beta$.

6. 设 V 是实内积空间, 并设 $T\in\mathcal{L}(V)$. 证明 V 有规范正交基使得 T 关于此基有分块上三角矩阵

$$\begin{bmatrix} \boldsymbol{A}_1 & & * \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中每个 $\boldsymbol{A}_j$ 都是 1×1 矩阵或没有本征值的 2×2 矩阵.

7. 证明: 如果 $T \in \mathcal{L}(V)$, 并且 j 是正整数使得 $j \leqslant \dim V$, 则 T 有 $j-1$ 维或 j 维的不变子空间.

8. 证明: 对于任何 $T \in \mathcal{L}(\mathbf{R}^7)$, 算子 T^2+T+I 都不是幂零的.

9. 给出一个算子 $T \in \mathcal{L}(\mathbf{C}^7)$ 使得 T^2+T+I 是幂零的.

10. 设 V 是实向量空间, $T \in \mathcal{L}(V)$, 并且 $\alpha, \beta \in \mathbf{R}$ 使得 $\alpha^2 < 4\beta$. 证明对于每个正整数 k,

$$\operatorname{null}(T^2+\alpha T+\beta I)^k$$

的维数都是偶数.

11. 设 V 是实向量空间, $T \in \mathcal{L}(V)$, 并设 $\alpha, \beta \in \mathbf{R}$, 使得 $\alpha^2 < 4\beta$ 并且 $T^2+\alpha T+\beta I$ 是幂零的. 证明 $\dim V$ 是偶数, 并且

$$(T^2+\alpha T+\beta I)^{\dim V/2} = 0.$$

12. 证明: 若 $T \in \mathcal{L}(\mathbf{R}^3)$, 并且 5 和 7 都是 T 的本征值, 则 T 没有本征对.

13. 设 V 是 n 维实向量空间, 并且 $T \in \mathcal{L}(V)$ 使得

$$\operatorname{null} T^{n-2} \neq \operatorname{null} T^{n-1}.$$

证明 T 最多有两个互不相同的本征值, 而且 T 没有本征对.

14. 设 V 是 2 维向量空间, 并设 $T \in \mathcal{L}(V)$. 证明: 如果

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

是 T 关于 V 的某个基的矩阵, 则 T 的特征多项式等于 $(z-a)(z-d)-bc$.

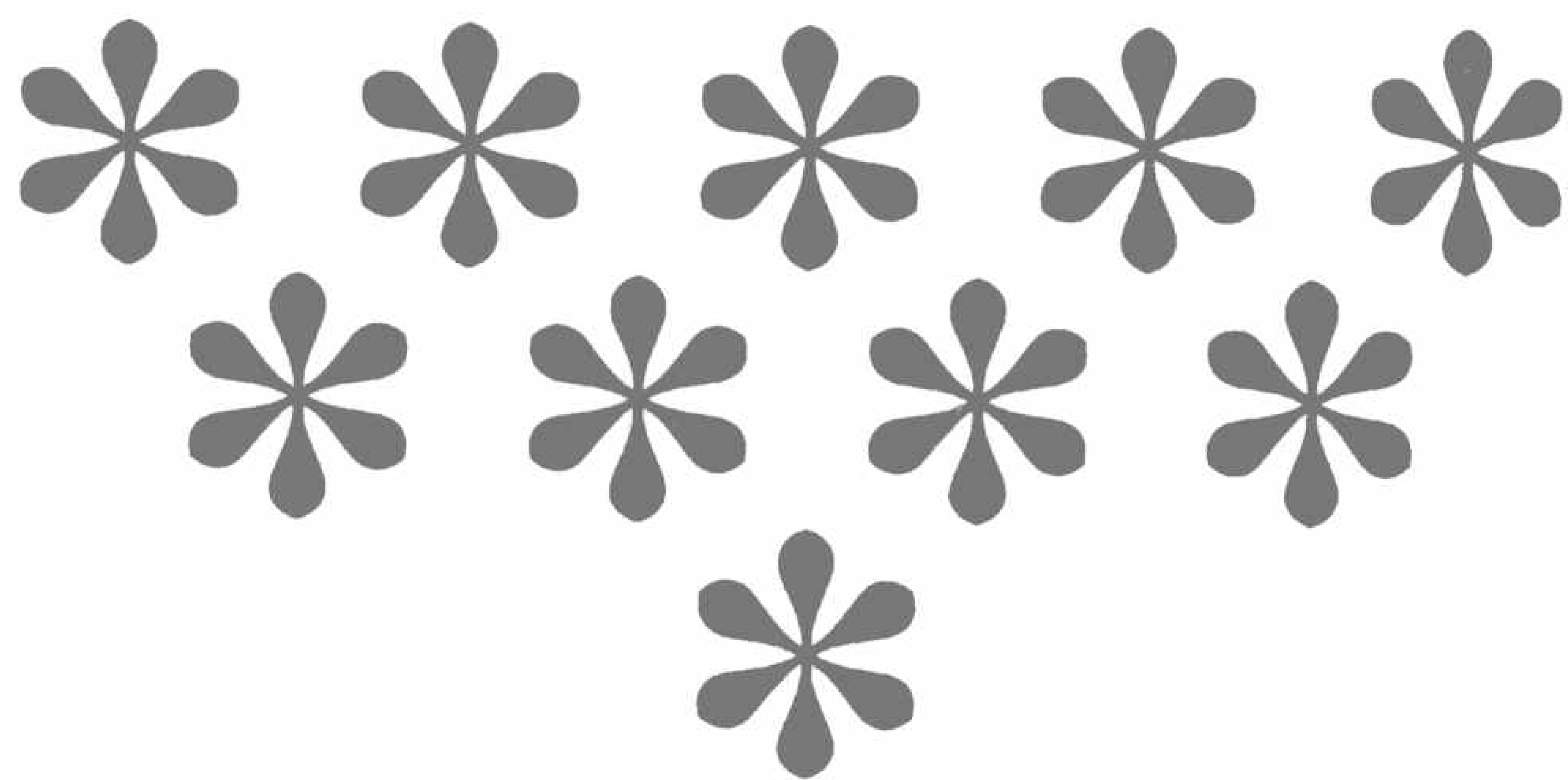
做这个习题无需求出 T 的本征值. 和平时一样, 除非特别指明, 这里的 V 是实向量空间或复向量空间.

15. 设 V 是实内积空间, 并设 $S \in \mathcal{L}(V)$ 是等距同构. 证明: 若 (α, β) 是 S 的本征对, 则 $\beta = 1$.

第 10 章　迹与行列式

本书的重点始终是线性映射和算子, 而不是矩阵. 本章将定义并讨论迹和行列式, 因此对矩阵的关注要多一些. 行列式只出现在本书的结尾, 这是因为我们用更自然的处理方法取代了行列式在线性代数中通常的应用 (定义特征多项式以及证明复向量空间上的算子都有本征值). 本书在最后说明了行列式在体积和积分理论中所起到的重要作用.

> $\mathbf{F}$ 表示 $\mathbf{R}$ 或 $\mathbf{C}$.
> V 是 $\mathbf{F}$ 上的有限维非零向量空间.



§10.1　基　变　换

算子 $T \in \mathcal{L}(V)$ 的矩阵依赖于 V 的基的选取. V 的两个不同的基可能给出 T 的不同的矩阵. 本节我们将讨论这些矩阵之间的联系. 这个信息有助于我们在本章的后面找到 T 的迹和行列式的公式.

恒等算子 $I \in \mathcal{L}(V)$ 关于 V 的任意基都有对角矩阵

$$\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}.$$

这个矩阵称为**单位矩阵 (identity matrix)**, 记为 $\boldsymbol{I}$. 注意, 我们用符号 $\boldsymbol{I}$ 表示 (所有向量空间上的) 恒等算子和 (所有可能大小的) 单位矩阵. 你应该能从上下文中确定 I 指的是什么. 例如, 考虑等式

$$\mathcal{M}(I) = \boldsymbol{I};$$

左端的 I 代表恒等算子, 右端的 $\boldsymbol{I}$ 代表单位矩阵.

如果 $\boldsymbol{A}$ 是一个与 $\boldsymbol{I}$ 大小相同的方阵 (像通常一样, 其中的元素属于 $\mathbf{F}$), 那么你应该验证 $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$. 方阵 $\boldsymbol{A}$ 称为**可逆的 (invertible)**, 如果存在一个同样大小的方阵 $\boldsymbol{B}$ 使得 $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$. 此时, 称 $\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的**逆 (inverse)**. 要证明 $\boldsymbol{A}$ 最多有一个逆, 设 $\boldsymbol{B}$ 和 $\boldsymbol{B}'$ 都是 $\boldsymbol{A}$ 的逆, 那么

有些数学家使用术语**非奇异的 (nonsingular)** 和**奇异的 (singular)**, 意思分别与可逆和不可逆相同.

$$\boldsymbol{B} = \boldsymbol{BI} = \boldsymbol{B}(\boldsymbol{AB}') = (\boldsymbol{BA})\boldsymbol{B}' = \boldsymbol{IB}' = \boldsymbol{B}',$$

因此 $\boldsymbol{B} = \boldsymbol{B}'$. 因为逆是唯一的, 所以 (如果 $\boldsymbol{A}$ 是可逆的) 可以用记号 $\boldsymbol{A}^{-1}$ 来表示 $\boldsymbol{A}$ 的逆. 也就是说, 如果 $\boldsymbol{A}$ 是可逆的, 那么 $\boldsymbol{A}^{-1}$ 是唯一一个与 $\boldsymbol{A}$ 大小相同且使得 $\boldsymbol{AA}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$ 的矩阵.

回想一下, 在第 3 章讨论从一个向量空间到另一个向量空间的线性映射时, 我们定义了一个线性映射关于两个基 —— 其中一个是第一个向量空间的基, 另一个是第二个向量空间的基的矩阵. 算子是一个向量空间到自身的线性映射, 我们在讨论算子时几乎总是对两个向量空间使用同一个基 (毕竟所讨论的两

个向量空间是相同的). 因此, 通常我们在提及一个算子关于某个基的矩阵时, 指的就是在这两个相同的向量空间中使用同一个基. 下一个命题是一种少见的情况, 即使对于向量空间到其自身的同一个算子, 我们也要用到两个不同的基.

回想一下矩阵乘法和线性映射的乘法之间的联系. 假设除了 V 之外我们还有另外两个有限维线性空间, 比如 U 和 W. 设 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p)$ 是 U 的基, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基, $(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m)$ 是 W 的基. 如果 $T \in \mathcal{L}(U, V)$, $S \in \mathcal{L}(V, W)$, 那么 $ST \in \mathcal{L}(U, W)$, 并且

10.1

$$\begin{aligned} &\mathcal{M}(ST, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p), (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m)) \\ &= \mathcal{M}(S, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n), (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m)) \\ &\quad \times \mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)). \end{aligned}$$

上式成立是根据矩阵乘法的定义 —— 参见 3.11 及其后面的内容.

下面的命题讨论的是恒等算子关于两个不同的基的矩阵. 注意, 把 $\boldsymbol{u}_k$ 写成这些 $\boldsymbol{v}$ 的线性组合, 所用的标量就构成 $\mathcal{M}(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))$ 的第 k 列. 作为下一个命题的例子, 考虑 $\mathbf{F}^2$ 的基 $((4,2),(5,3))$ 和 $((1,0),(0,1))$. 显然

$$\mathcal{M}(I, ((4,2),(5,3)), ((1,0),(0,1))) = \begin{bmatrix} 4 & 5 \\ 2 & 3 \end{bmatrix}.$$

你应该验证, 上面矩阵的逆是 $\begin{bmatrix} 3/2 & -5/2 \\ -1 & 2 \end{bmatrix}$. 于是由下面的命题可得

$$\mathcal{M}(I, ((1,0),(0,1)), ((4,2),(5,3))) = \begin{bmatrix} 3/2 & -5/2 \\ -1 & 2 \end{bmatrix}$$

10.2 命题: *如果 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ 和 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 都是 V 的基, 那么 $\mathcal{M}(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))$ 是可逆的, 并且*

$$\begin{aligned} &\mathcal{M}(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))^{-1} \\ &= \mathcal{M}(I, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n), (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)). \end{aligned}$$

证明: 在 10.1 中用 V 代替 U 和 W, 用 $\boldsymbol{u}_j$ 代替 $\boldsymbol{w}_j$, 并且用 I 代替 S 和 T, 得到

$$I = \mathcal{M}(I, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n), (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)) \\ \times \mathcal{M}(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)).$$

再互换诸 $\boldsymbol{u}$ 和诸 $\boldsymbol{v}$ 的角色, 得到

$$I = \mathcal{M}(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)) \\ \times \mathcal{M}(I, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n), (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)).$$

由这两个等式即得所要的结果. ■

现在我们可以看到在改变基时 T 的矩阵是怎样变化的.

10.3 定理: 假设 $T \in \mathcal{L}(V)$, $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ 和 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 都是 V 的基, 并设 $\boldsymbol{A} = \mathcal{M}\big(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)\big)$, 那么

10.4 $$\mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)) = \boldsymbol{A}^{-1}\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))\boldsymbol{A}.$$

证明: 在 10.1 中用 V 代替 U 和 W, 用 $\boldsymbol{v}_j$ 代替 $\boldsymbol{w}_j$, 用 I 代替 T, 用 T 代替 S, 得到

10.5 $$\mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)) = \mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))\boldsymbol{A}.$$

再利用 10.1, 此时用 V 代替 U 和 W, 用 $\boldsymbol{u}_j$ 代替 $\boldsymbol{w}_j$, 并且用 I 代替 S, 得到

$$\mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)) = \boldsymbol{A}^{-1}\mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)).$$

这里用到了 10.2. 将 10.5 带入上面的等式即得 10.4. ■

§10.2 迹

在前两章我们已经研究了特征多项式, 本章将进一步研究特征多项式. 如果 V 是一个 n 维复向量空间, 并且 $T \in \mathcal{L}(V)$, 那么 T 的特征多项式等于

$$(z - \lambda_1) \cdots (z - \lambda_n),$$

其中 $\lambda_1, \cdots, \lambda_n$ 是 T 的所有本征值, 按照重数重复. 将上面的多项式展开, 我们可以将 T 的特征多项式写成如下形式

10.6 $$z^n - (\lambda_1 + \cdots + \lambda_n)z^{n-1} + \cdots + (-1)^n(\lambda_1 \cdots \lambda_n).$$

如果 V 是 n 维实向量空间, 并且 $T \in \mathcal{L}(V)$, 那么 T 的特征多项式等于

$$(x-\lambda_1)\cdots(x-\lambda_m)(x^2+\alpha_1 x+\beta_1)\cdots(x^2+\alpha_M x+\beta_M),$$

这里 m 和 M 都可能等于 0.

其中 $\lambda_1, \cdots, \lambda_m$ 是 T 的所有本征值, $(\alpha_1, \beta_1), \cdots, (\alpha_M, \beta_M)$ 是 T 的所有本征对, 每个都按照重数重复. 将上面的多项式展开, 我们可以将 T 的特征多项式写成如下形式

回想一下, 实数对 (α, β) 是 T 的本征对, 如果 $\alpha^2 < 4\beta$, 并且 $T^2+\alpha T+\beta I$ 不是单的.

10.7 $$x^n-(\lambda_1+\cdots+\lambda_m-\alpha_1-\cdots-\alpha_M)x^{n-1}+\cdots \\ +(-1)^m(\lambda_1\cdots\lambda_m\beta_1\cdots\beta_M).$$

本节将讨论特征多项式中 z^{n-1} (在讨论实向量空间时, 通常记为 x^{n-1}) 的系数. 下一节将讨论特征多项式中的常数项.

对于 $T \in \mathcal{L}(V)$, T 的特征多项式中 z^{n-1} (或者对于实向量空间, x^{n-1}) 的系数的相反数称为 T 的**迹 (trace)**, 记为 $\operatorname{trace} T$. 如果 V 是复向量空间, 那么 10.6 表明 T 的迹等于 T 的本征值之和, 计重数. 如果 V 是实向量空间, 那么 10.7 表明 T 的迹等于 T 的本征值之和减去 T 的本征对的第一个坐标之和, 每个都按照重数重复. 例如, 设算子 $T \in \mathcal{L}(\mathbf{C}^3)$ 的矩阵为

注意, $\operatorname{trace} T$ 只依赖于 T, 而不依赖于 V 的基, 这是因为 T 的特征多项式不依赖于基的选取.

10.8 $$\begin{bmatrix} 3 & -1 & -2 \\ 3 & 2 & -3 \\ 1 & 2 & 0 \end{bmatrix}.$$

于是, 你可以验证, T 的本征值是 $1, 2+3i, 2-3i$, 重数都为 1. 计算本征值之和, 得到 $\operatorname{trace} T = 1+(2+3i)+(2-3i)$; 也就是说, $\operatorname{trace} T = 5$.

又如, 设 $T \in \mathcal{L}(\mathbf{R}^3)$ 也是一个矩阵为 10.8 的算子 (注意, 上一段讨论的是复向量空间; 现在讨论的是实向量空间). 那么, 在上一章中 (参见 205 页) 你应该已经验证了: 1 是 T 的唯一本征值 (重数为 1), 并且 $(-4, 13)$ 是 T 的唯一本征对 (重数为 1). 计算本征值之和减去本征对的第一个坐标之和, 得到 $\operatorname{trace} T = 1-(-4)$, 即 $\operatorname{trace} T = 5$.

在找到利用算子的矩阵来计算迹的公式 (对复向量空间和实向量空间都适用) 以后, 前面两个例子中的算子的迹相同的原因就清楚了.

本节的其余部分主要讨论如何利用 T (关于任意一个基) 的矩阵来计算 T 的迹. 首先考虑最简单的情形. 设 V 是复向量空间, $T \in \mathcal{L}(V)$, 取 V 的一个基使得 T 关于此基有上三角矩阵 $\boldsymbol{A}$. 于是 T 的本征值恰好是 $\boldsymbol{A}$ 的对角线元素, 按照重数重复 (参见 8.10). 于是 $\operatorname{trace} T$ 等于 $\boldsymbol{A}$ 的对角线元素之和. 同样的公式也适用于矩阵为 10.8 并且迹等于 5 的那个算子 $T \in \mathcal{L}(\mathbf{F}^3)$. 如此简单的公式普遍成立吗?

首先研究 $T \in \mathcal{L}(V)$, 其中 V 是实向量空间. 取 V 的一个基使得 T 关于此基有分块上三角矩阵 $\mathcal{M}(T)$, 其中对角线上的每个块都是由 T 的本征值组成的 $1{\times}1$ 块或者是没有本征值的 $2{\times}2$ 块 (参见 9.4 和 9.9). $\mathcal{M}(T)$ 的对角线上的每个 $1{\times}1$ 块的元素都是 T 的本征值, 因此对 $\operatorname{trace} T$ 的值是有贡献的. 如果 $\mathcal{M}(T)$ 的对角线上有 $2{\times}2$ 块, 那么考虑其中一个典型的块

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

这个 $2{\times}2$ 矩阵的特征多项式是 $(x-a)(x-d)-bc$, 即

$$x^2-(a+d)x+(ad-bc).$$

你应该认真复习 9.9, 只有这样才能理解本征对与 2×2 块的特征多项式的关系.

因此 $(-a-d, ad-bc)$ 是 T 的本征对. 这个本征对第一个坐标的相反数 $a+d$ 就是这个块对 $\operatorname{trace} T$ 的值的贡献. 注意, $a+d$ 是这个块中对角线元素之和. 于是, 对于 V 的一个基, 如果 T 关于此基具有 9.4 或 9.9 所要求的分块上三角矩阵, 那么 $\operatorname{trace} T$ 等于对角线元素之和.

现在你可能会猜到, $\operatorname{trace} T$ 等于 T 关于任意基的矩阵的对角线元素之和. 可以证明确实如此. 为此, 定义方阵 $\boldsymbol{A}$ 的**迹 (trace)** 为其对角线元素之和, 记为 $\operatorname{trace} \boldsymbol{A}$. 采用这样的记号, 我们将证明 $\operatorname{trace} T = \operatorname{trace} \mathcal{M}(T,(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n))$, 其中 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 是 V 的任意一个基. 我们已经知道, 如果 T 关于 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 具有上三角矩阵 (若 V 是复的) 或者具有适当的分块上三角矩

阵 (若 V 是实的), 那么这个结论成立. 为了对任意基证明这个迹公式, 我们需要下面的命题.

10.9 命题: 如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是大小相同的方阵, 那么

$$\operatorname{trace}(\boldsymbol{AB}) = \operatorname{trace}(\boldsymbol{BA}).$$

证明: 设

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & & \vdots \\ b_{n,1} & \cdots & b_{n,n} \end{bmatrix}.$$

$\boldsymbol{AB}$ 的对角线上的第 j 项等于

$$\sum_{k=1}^{n} a_{j,k} b_{k,j}.$$

因此

$$\begin{aligned} \operatorname{trace}(\boldsymbol{AB}) &= \sum_{j=1}^{n}\sum_{k=1}^{n} a_{j,k} b_{k,j} \\ &= \sum_{k=1}^{n}\sum_{j=1}^{n} b_{k,j} a_{j,k} \\ &= \sum_{k=1}^{n} \boldsymbol{BA} \text{ 对角线上的第 } k \text{ 个元素} \\ &= \operatorname{trace}(\boldsymbol{BA}). \end{aligned}$$

■

现在可以证明, 算子关于一个基的矩阵的对角线元素之和并不依赖于这个基的选取.

10.10 推论: 设 $T \in \mathcal{L}(V)$. 如果 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ 和 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 都是 V 的基, 那么

$$\operatorname{trace}\mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)) = \operatorname{trace}\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)).$$

证明：假设 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ 和 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 都是 V 的基，并设

$$\boldsymbol{A} = \mathcal{M}\big(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)\big),$$

那么

这里的第三个等式依赖于矩阵乘法的结合性质.

$$\begin{aligned}\operatorname{trace}\mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)) &= \operatorname{trace}(\boldsymbol{A}^{-1}(\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))\boldsymbol{A}))\\ &= \operatorname{trace}\left((\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))\boldsymbol{A})\boldsymbol{A}^{-1}\right)\\ &= \operatorname{trace}\mathcal{M}\left(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)\right),\end{aligned}$$

其中第一个等式由 10.3 得到，第二个等式由 10.9 得到. 第三个等式完成了证明. ■

下面的定理指出，算子的迹等于该算子的矩阵的对角线元素之和. 这个定理没有指明所用到的基，这是因为根据上面的推论，对于每个基来说，算子的矩阵的对角线元素之和都相同.

10.11 定理：若 $T \in \mathcal{L}(V)$，则 $\operatorname{trace} T = \operatorname{trace}\mathcal{M}(T)$.

证明：令 $T \in \mathcal{L}(V)$. 上面已经注意到，$\operatorname{trace}\mathcal{M}(T)$ 与 V 的基的选取无关 (由 10.10). 因此，要证明对每个基都有

$$\operatorname{trace} T = \operatorname{trace}\mathcal{M}(T),$$

只需证明上式对 V 的某个基成立. 取 V 的一个基使得关于此基 $\mathcal{M}(T)$ 是上三角矩阵 (如果 V 是复向量空间) 或者是适当的分块上三角矩阵 (如果 V 是实向量空间)，我们已经证明了 (在 218 页) 关于这个基上式成立. ■

如果知道了复向量空间上一个算子的矩阵，利用上面的定理，不求算子的任何本征值就可以求出所有本征值的和. 例如，考虑 $\mathbf{C}^5$ 上的一个算子，它的矩阵如下

$$\begin{bmatrix}0&0&0&0&-3\\1&0&0&0&6\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\end{bmatrix}.$$

没人知道这个算子的任何本征值的精确公式，但是我们却知道

它的本征值之和等于 0, 这是因为上面矩阵的对角线元素之和等于 0.

通过转换成矩阵的迹的语言, 利用上面的定理容易证明算子的迹的一些有用性质, 其中有些性质是已经证明过的, 或者是显然的. 在下一个推论的证明中, 我们就是这么做的.

10.12 推论: 若 $S, T \in \mathcal{L}(V)$, 则

$$\operatorname{trace}(ST) = \operatorname{trace}(TS),$$

$$\operatorname{trace}(S+T) = \operatorname{trace} S + \operatorname{trace} T.$$

证明: 设 $S, T \in \mathcal{L}(V)$. 任取 V 的一个基, 则

$$\begin{aligned}
\operatorname{trace}(ST) &= \operatorname{trace}\mathcal{M}(ST) \\
&= \operatorname{trace}\left(\mathcal{M}(S)\mathcal{M}(T)\right) \\
&= \operatorname{trace}\left(\mathcal{M}(T)\mathcal{M}(S)\right) \\
&= \operatorname{trace}\mathcal{M}(TS) \\
&= \operatorname{trace}(TS),
\end{aligned}$$

其中第一个等式和最后一个等式由 10.11 得到, 中间的等式由 10.9 得到. 这就完成了第一个结论的证明.

要证明第二个结论, 注意到

$$\begin{aligned}
\operatorname{trace}(S+T) &= \operatorname{trace}\mathcal{M}(S+T) \\
&= \operatorname{trace}\left(\mathcal{M}(S)+\mathcal{M}(T)\right) \\
&= \operatorname{trace}\mathcal{M}(S) + \operatorname{trace}\mathcal{M}(T) \\
&= \operatorname{trace} S + \operatorname{trace} T,
\end{aligned}$$

其中第一个等式和最后一个等式也是由 10.11 得到的; 根据矩阵的迹的定义, 第三个等式是显然的. 这就完成了第二个结论的证明. ■

利用前面这些方法可以得到下面的奇妙推论. 这个结果在无限维向量空间上的推广可以导出量子理论的一些重要结论.

这个推论的叙述并未涉及迹，但其简短的证明却用到了迹. 在数学中一旦有类似的事情发生，我们可以确信背后一定隐藏着一个很好的定义.

10.13 推论: 不存在算子 $S,T\in\mathcal{L}(V)$ 使得 $ST-TS=I$.

证明: 设 $S,T\in\mathcal{L}(V)$, 那么

$$\begin{aligned}\operatorname{trace}(ST-TS)&=\operatorname{trace}(ST)-\operatorname{trace}(TS)\\&=0,\end{aligned}$$

其中第二个等式由 10.12 得到. 显然 I 的迹等于 $\dim V$, 不等于 0. 因为 $ST-TS$ 和 I 有不同的迹, 所以它们不相等. ■

§10.3 算子的行列式

注意, $\det T$ 只依赖于 T 而不依赖于 V 的基, 因为 T 的特征多项式不依赖于基的选取.

对于 $T\in\mathcal{L}(V)$, 定义 T 的**行列式 (determinant)** 为 T 的特征多项式的常数项乘以 $(-1)^{\dim V}$, 记为 $\det T$. 定义中使用因子 $(-1)^{\dim V}$ 的动机来自 10.6.

如果 V 是复向量空间, 那么由 10.6 立即可得, $\det T$ 等于 T 的本征值的乘积, 计重数. 回想一下, 如果 V 是复向量空间, 那么 V 有基使得 T 关于此基具有上三角矩阵 (参见 5.13); 因此 $\det T$ 等于这个矩阵的对角线元素之积 (参见 8.10).

如果 V 是实向量空间, 那么 $\det T$ 等于 T 的本征值之积再乘以 T 的本征对的第二个坐标之积, 每个都按照重数重复——这里使用了 10.7 并观察到 $m=\dim V-2M$ (使用 10.7 的记号), 于是 $(-1)^m=(-1)^{\dim V}$.

例如, 设算子 $T\in\mathcal{L}(\mathbf{C}^3)$ 的矩阵为 10.8. 在上一节我们知道, T 的本征值为 $1, 2+3i, 2-3i$, 并且重数都为 1. 计算本征值的乘积得 $\det T=(1)(2+3i)(2-3i)$, 即 $\det T=13$.

又如, 设算子 $T\in\mathcal{L}(\mathbf{R}^3)$ 的矩阵为 10.8 (注意, 上一段讨论的是复向量空间; 现在讨论的是实向量空间), 我们早就注意到了, 1 是 T 的唯一本征值 (重数为 1), 并且 $(-4,13)$ 是 T 的唯一本征对 (重数为 1). 计算本征值之积与本征对的第二个坐标之积的乘积得 $\det T=(1)(13)$, 即 $\det T=13$.

当找到利用算子的矩阵来计算其行列式的公式 (对复向量空间和实向量空间都适用) 以后, 前面两个例子中的算子的行列式相同的原因就清楚了.

本节将证明行列式的一些简单而重要的性质. 下一节将会发现利用 T (关于任意基) 的矩阵计算 $\det T$ 的方法. 首先来证明一个至关重要的结果, 我们的处理方法使得它的证明非常简单.

10.14 命题: 一个算子是可逆的当且仅当它的行列式不等于零.

证明: 首先, 设 V 是复向量空间, 并且 $T \in \mathcal{L}(V)$. 算子 T 是可逆的当且仅当 0 不是 T 的本征值. 显然, 这个条件成立当且仅当 T 的本征值的乘积不等于 0. 因此, T 是可逆的当且仅当 $\det T \neq 0$.

现在假设 V 是实向量空间, 并且 $T \in \mathcal{L}(V)$. 此时仍有, T 是可逆的当且仅当 0 不是 T 的本征值. 采用 10.7 的记号有

10.15 $$\det T = \lambda_1 \cdots \lambda_m \beta_1 \cdots \beta_M,$$

其中这些 λ 是 T 的所有本征值, 这些 β 是 T 的所有本征对的第二个坐标, 每个都按照重数重复. 对于每个本征对 (α_j, β_j) 都有 $\alpha_j^2 < 4\beta_j$. 特别地, 每个 β_j 都是正的. 从而 (参见 10.15), $\lambda_1 \cdots \lambda_m \neq 0$ 当且仅当 $\det T \neq 0$. 于是 T 是可逆的当且仅当 $\det T \neq 0$. ■

如果 $T \in \mathcal{L}(V)$, 并且 $\lambda, z \in \mathbf{F}$, 那么 λ 是 T 的本征值当且仅当 $z - \lambda$ 是 $zI - T$ 的本征值, 这是因为

$$-(T - \lambda I) = (zI - T) - (z - \lambda)I.$$

等式两端同时取 $\dim V$ 次幂, 然后再取零空间可知, λ 作为 T 的本征值的重数等于 $z - \lambda$ 作为 $zI - T$ 的本征值的重数. 下一个引理给出了关于本征对的一个类似的结果. 我们将利用这个引

理来证明特征多项式可以表示成某个行列式.

实向量空间比复向量空间更难处理. 你第一次读这一章时, 可以只考虑复向量空间, 把精力集中于基本思想, 而忽略处理实向量空间所需要的特殊方法.

10.16 引理: 设 V 是实向量空间, $T \in \mathcal{L}(V)$, $\alpha, \beta, x \in \mathbf{R}$, 并且 $\alpha^2 < 4\beta$, 那么 (α, β) 是 T 的本征对当且仅当 $(-2x-\alpha, x^2+\alpha x+\beta)$ 是 $xI-T$ 的本征对. 进一步, 这两个本征对的重数相同.

证明: 首先, 需要验证 $(-2x-\alpha, x^2+\alpha x+\beta)$ 满足本征对所要求的不等式. 我们有

$$\begin{aligned}(-2x-\alpha)^2 &= 4x^2+4\alpha x+\alpha^2\\ &< 4x^2+4\alpha x+4\beta\\ &= 4(x^2+\alpha x+\beta).\end{aligned}$$

因此 $(-2x-\alpha, x^2+\alpha x+\beta)$ 满足所需要的不等式.

现在, 你应该验证

$$T^2+\alpha T+\beta I = (xI-T)^2-(2x+\alpha)(xI-T)+(x^2+\alpha x+\beta)I,$$

于是, (α, β) 是 T 的本征对当且仅当 $(-2x-\alpha, x^2+\alpha x+\beta)$ 是 $xI-T$ 的本征对. 进一步, 等式两端取 $\dim V$ 次幂, 然后再取零空间即知重数相等. ■

大部分教材都把下面的定理作为特征多项式的定义. 采用这种处理方法的教材必需在接触有趣的线性代数之前用大量的时间来发展行列式理论.

10.17 定理: 设 $T \in \mathcal{L}(V)$, 则 T 的特征多项式等于 $\det(zI-T)$.

证明: 首先, 设 V 是复向量空间. 令 $\lambda_1, \cdots, \lambda_n$ 表示 T 的本征值, 按照重数重复. 于是对 $z \in \mathbf{C}$, $zI-T$ 的本征值为 $z-\lambda_1, \cdots, z-\lambda_n$, 按照重数重复. $zI-T$ 的行列式是这些本征值的乘积. 也就是说,

$$\det(zI-T) = (z-\lambda_1)\cdots(z-\lambda_n).$$

根据定义, 上面等式的右端是 T 的特征多项式, 这就完成了 V

是复向量空间的情形的证明.

现在假设 V 是实向量空间. 令 $\lambda_1, \cdots, \lambda_m$ 表示 T 的所有本征值, $(\alpha_1, \beta_1), \cdots, (\alpha_M, \beta_M)$ 表示 T 的所有本征对, 每个都按照重数重复. 于是对 $x \in \mathbf{R}$, $xI - T$ 的本征值为 $x - \lambda_1, \cdots, x - \lambda_m$, 并且由 10.16 知 $xI - T$ 的本征对为

$$(-2x - \alpha_1, x^2 + \alpha_1 x + \beta_1), \cdots, (-2x - \alpha_M, x^2 + \alpha_M x + \beta_M),$$

每个都按照重数重复. 因此

$$\begin{aligned}&\det(xI - T)\\ &= (x - \lambda_1) \cdots (x - \lambda_m)(x^2 + \alpha_1 x + \beta_1) \cdots (x^2 + \alpha_M x + \beta_M).\end{aligned}$$

根据定义, 上面等式的右端是 T 特征多项式, 这就完成了 V 是实向量空间情形的证明. ■

§10.4 矩阵的行列式

本节主要讨论如何利用 T (关于任意基) 的矩阵来计算 $\det T$. 首先讨论最简单的情形. 设 V 是复向量空间, $T \in \mathcal{L}(V)$, 并取 V 的基使得 T 关于此基有上三角矩阵. 在上一节我们注意到, $\det T$ 等于这个矩阵的对角线元素之积. 这样简单的公式普遍成立吗?

可惜行列式比迹要复杂得多. 特别地, $\det T$ 未必等于 T 关于任何基的矩阵 $\mathcal{M}(T)$ 的对角线元素之积. 例如, 上一节我们看到, $\mathbf{F}^3$ 上具有矩阵 10.8 的算子的行列式等于 13, 但是那个矩阵的对角线元素之积却等于 0.

对于每个方阵 $\boldsymbol{A}$, 我们都要定义 $\boldsymbol{A}$ 的行列式, 记为 $\det \boldsymbol{A}$, 使得无论用哪个基来计算 $\mathcal{M}(T)$ 都有 $\det T = \det \mathcal{M}(T)$. 为了寻找矩阵行列式的正确定义, 先来计算某些特殊算子的行列式.

设 $c_1, \cdots, c_n \in \mathbf{F}$ 都是非零标量, 并且 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基. 考虑算子 $T \in \mathcal{L}(V)$ 使得 $\mathcal{M}\big(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)\big)$ 等于

10.18
$$\begin{bmatrix} 0 & & & & c_n \\ c_1 & 0 & & & \\ & c_2 & 0 & & \\ & & \ddots & \ddots & \\ & & & c_{n-1} & 0 \end{bmatrix};$$

在这个矩阵中, 除了右上角的元素和紧位于对角线之下的那条直线上的元素之外, 其余元素都等于 0. 我们来求 T 的行列式. 注意到

$$\begin{aligned} &(\boldsymbol{v}_1, T\boldsymbol{v}_1, T^2\boldsymbol{v}_1, \cdots, T^{n-1}\boldsymbol{v}_1) \\ &= (\boldsymbol{v}_1, c_1\boldsymbol{v}_2, c_1c_2\boldsymbol{v}_3, \cdots, c_1\cdots c_{n-1}\boldsymbol{v}_n). \end{aligned}$$

因此 $(\boldsymbol{v}_1, T\boldsymbol{v}_1, \cdots, T^{n-1}\boldsymbol{v}_1)$ 是线性无关的 (所有的 c 都非零). 于是, 如果 p 是次数不超过 $n-1$ 的非零多项式, 那么 $p(T)\boldsymbol{v}_1 \neq \mathbf{0}$. 也就是说, T 的极小多项式的次数不可能小于 n. 你应该验证, 对每个 j 都有 $T^n\boldsymbol{v}_j = c_1\cdots c_n\boldsymbol{v}_j$, 于是 $T^n = c_1\cdots c_nI$. 因此 $z^n - c_1\cdots c_n$ 是 T 的极小多项式. 因为 $n = \dim V$, 所以 $z^n - c_1\cdots c_n$ 也是 T 的特征多项式. 用 $(-1)^n$ 乘以这个多项式的常数项得

回想一下, 如果算子 $T \in \mathcal{L}(V)$ 的极小多项式的次数为 $\dim V$, 那么 T 的特征多项式等于 T 的极小多项式. 计算极小多项式通常是求特征多项式的有效的方法.

10.19
$$\det T = (-1)^{n-1}c_1\cdots c_n.$$

如果某个 c_j 等于 0, 那么显然 T 不是可逆的, 从而 $\det T = 0$, 于是同样的公式仍然成立. 因此, 为使 $\det T = \det \mathcal{M}(T)$, 我们必须让 10.18 的行列式等于 $(-1)^{n-1}c_1\cdots c_n$. 但我们现在还没有足够的证据来给出任意方阵的行列式定义的一个合理的猜想.

为了计算一类更复杂的算子的行列式, 我们引入排列的概念. $(1, \cdots, n)$ 的一个**排列 (permutation)** 是一个组 $(m_1, \cdots, m_n)$, 其中 $1, \cdots, n$ 中的每个数恰好出现一次. $(1, \cdots, n)$ 的所有排列组成的集合记为 $\operatorname{perm} n$. 例如, $(2, 3, \cdots, n, 1) \in \operatorname{perm} n$. 你应该想到, 把前 n 个正整数按某种顺序排起来就是 $\operatorname{perm} n$ 的一个元素. 为了简单起见, 我们将讨论由复数组成的矩阵 (现阶段我们只提供动机 —— 正式的证明将在后面给出). 假设 V 是复向量空间, 并设 $c_1, \cdots, c_n \in \mathbf{C}$, $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基. 考虑可以如下得到的排列 $(p_1, \cdots, p_n) \in \operatorname{perm} n$: 将 $(1, \cdots, n)$ 拆成

连续整数的组, 然后把每组的第一项移到该组的最后. 例如, 取 $n=9$, 则排列

10.20 $$(2,3,1,5,6,7,4,9,8)$$

可以这样得到: 将 $(1,2,3),(4,5,6,7),(8,9)$ 各组中的第一项移到最后, 得到 $(2,3,1),(5,6,7,4),(9,8)$, 然后再把它们放在一起即得 10.20. 设算子 $T\in\mathcal{L}(V)$ 使得

10.21 $$T\boldsymbol{v}_k=c_k\boldsymbol{v}_{p_k},\quad k=1,\cdots,n.$$

我们想求 $\det T$ 的公式. 这推广了前面的例子, 这是因为, 如果 $(p_1,\cdots,p_n)$ 恰好是排列 $(2,3,\cdots,n,1)$, 那么矩阵等于 10.18 的算子 T 与 10.21 所定义的算子是相同的.

由 10.21 所定义的算子 T 关于基 $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n)$ 的矩阵是如下的分块对角矩阵

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_1 & & 0\\ & \ddots & \\ 0 & & \boldsymbol{A}_M\end{bmatrix},$$

其中每个块都是形如 10.18 的方阵. T 的所有本征值等于 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_M$ 的所有本征值的并集 (参见第 9 章习题 3). 回想一下, 复向量空间上的算子的行列式是其本征值的乘积, 可见, 方阵行列式的定义应该使得

$$\det\boldsymbol{A}=(\det\boldsymbol{A}_1)\cdots(\det\boldsymbol{A}_M).$$

然而我们已经知道如何计算每个与 10.18 同形 (当然 n 的取值可能不同) 的 $\boldsymbol{A}_j$ 的行列式. 把所有这些合起来应该有

$$\det\boldsymbol{A}=(-1)^{n_1-1}\cdots(-1)^{n_M-1}c_1\cdots c_n,$$

其中 $\boldsymbol{A}_j$ 的大小为 $n_j\times n_j$. 数 $(-1)^{n_1-1}\cdots(-1)^{n_M-1}$ 称为排列 $(p_1,\cdots,p_n)$ 的符号, 记为 $\operatorname{sign}(p_1,\cdots,p_n)$ (这只是一个临时的定义, 等到我们定义了任意排列的符号之后将改用一个等价的定义).

要把它写成不依赖于特殊排列 $(p_1, \cdots, p_n)$ 的形式, 令 $a_{j,k}$ 表示 $\boldsymbol{A}$ 中第 j 行第 k 列元素; 那么

$$a_{j,k} = \begin{cases} 0 & \text{若 } j \neq p_k; \\ c_k & \text{若 } j = p_k. \end{cases}$$

于是

10.22

$$\det \boldsymbol{A} = \sum_{(m_1, \cdots, m_n) \in \text{perm } n} \big(\text{sign}(m_1, \cdots, m_n)\big) a_{m_1,1} \cdots a_{m_n,n},$$

这是因为除了对应于排列 $(p_1, \cdots, p_n)$ 的那个和项之外, 其余的每个和项都等于 0.

现在考虑任意矩阵 $\boldsymbol{A}$, 其中第 j 行第 k 列的元素为 $a_{j,k}$. 受上一段的启发, 我们猜想 $\det \boldsymbol{A}$ 应该定义由 10.22 定义. 可以证明这是对的. 现在我们可以忽略动机, 开始更加正式的推导. 首先, 需要定义任意排列的符号.

有些教材使用不必要的奇特术语正负号函数 (signum), 它的意思就是符号.

如果在组 $(m_1, \cdots, m_n)$ 中, 使得 j 出现在 k 后面的整数对 (j, k), $1 \leqslant j < k \leqslant n$, 的个数是偶数, 则定义排列 $(m_1, \cdots, m_n)$ 的**符号 (sign)** 为 1; 如果这样的整数对的个数是奇数, 则定义排列 $(m_1, \cdots, m_n)$ 的符号为 -1. 也就是说, 如果自然顺序被改变了偶数次, 那么排列的符号等于 1; 如果自然顺序被改变了奇数次, 那么排列的符号等于 -1. 例如, 在排列 $(2, 3, \cdots, n, 1)$ 中, 使得 $j < k$ 并反序出现的整数对 (j, k) 只有 $(1, 2), (1, 3), \cdots, (1, n)$; 因为这样的对共有 $n - 1$ 个, 所以这个排列的符号等于 $(-1)^{n-1}$ (注意, 与 10.19 中出现的量一致).

通过将排列 $(1, 2, 3, 4)$ 中的前两个元素互换所得到的排列 $(2, 1, 3, 4)$ 的符号为 -1. 下一个引理表明, 互换任给排列中的任意两个元素都会改变该排列的符号.

10.23 引理: *互换排列中的两个元素将改变排列的符号.*

证明: 假设有两个排列, 其中第二个排列是通过互换第一个排列中的两个元素得到的. 如果被互换的两个元素在第一个排列中是以自然顺序出现的, 那么它们在第二个排列中就不再以自然顺序出现, 反之亦然. 因此, 互换使得这两个元素组成的数对的反序数目的改变量为 1 或者 -1 (两者都是奇数).

考虑介于被互换的两个元素之间的元素. 如果中间的一个元素最初和被互换的第一个元素是按照自然顺序出现的, 那么在互换之后它们就是反序出现的, 反之亦然. 类似地, 如果中间的一个元素最初和被互换的第二个元素是按照自然顺序出现的, 那么在互换之后它们就是反序出现的, 反之亦然. 于是, 对于中间的每个元素来说, 这样的互换最终使得与之有关的反序数对的数目的改变量为 2, 0, 或者 -2 (都是偶数).

对于所有其他的元素, 相关的反序数对的数量没有变化. 所以反序数对的数量的实际变化总量是一个奇数. 于是第二个排列的符号等于 -1 乘以第一个排列的符号. ■

如果 $\boldsymbol{A}$ 是 $n \times n$ 矩阵,

10.24
$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix},$$

那么 $\boldsymbol{A}$ 的**行列式 (determinant)**, 记为 $\det \boldsymbol{A}$, 定义为

这个定义的动机来自 10.22.

10.25
$$\det \boldsymbol{A} = \sum_{(m_1,\cdots,m_n)\in \operatorname{perm} n} \big(\operatorname{sign}(m_1,\cdots,m_n)\big) a_{m_1,1} \cdots a_{m_n,n}.$$

例如, 若 $\boldsymbol{A}$ 是 1×1 矩阵 $[a_{1,1}]$, 那么 $\det \boldsymbol{A} = a_{1,1}$, 这是因为 $\operatorname{perm} 1$ 中只有一个元素, 即 (1), 所以它的符号是 1. 再看一个更有趣的例子, 考虑一个典型的 2×2 矩阵. 显然 $\operatorname{perm} 2$ 含有两个元素, 即 $(1, 2)$ 和 $(2, 1)$, 它们的符号分别是 1 和 -1. 因此

10.26
$$\det \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}a_{2,2} - a_{2,1}a_{1,2}.$$

集合 $\operatorname{perm} 3$ 中含有 6 个元素. 一般来说, $\operatorname{perm} n$ 中含有 $n!$ 个元素. 注意, 当 n 增大时, $n!$ 迅速增大.

为了保证你理解这个过程, 你应该找出 3×3 矩阵

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$$

的行列式公式, 这只要利用上面给出的定义即可 (即使你已经知道答案也应该这么做).

我们来计算下面的上三角矩阵的行列式,

$$A = \begin{bmatrix} a_{1,1} & & * \\ & \ddots & \\ 0 & & a_{n,n} \end{bmatrix}.$$

排列 $(1,2,\cdots,n)$ 的符号为 1, 因此给出 $\det A$ 的定义 10.25 中的一项 $a_{1,1}\cdots a_{n,n}$. 任意其他的排列 $(m_1,\cdots,m_n)\in \operatorname{perm} n$ 至少包含一个元素 m_j 使得 $m_j > j$, 即 $a_{m_j,j} = 0$ (因为 A 是上三角矩阵). 于是 $\det A$ 的定义 10.25 中其他的每个和项对求和都没有贡献. 因此 $\det A = a_{1,1}\cdots a_{n,n}$. 也就是说, 上三角矩阵的行列式等于对角线元素的乘积. 特别地, 由此可得, 如果 V 是复向量空间, $T\in\mathcal{L}(V)$, 并且取 V 的一个基使得关于此基 $\mathcal{M}(T)$ 是上三角矩阵, 那么 $\det T = \det\mathcal{M}(T)$. 我们的目标是证明这个结论对 V 的每个基都成立, 而不只是对给出上三角矩阵的基成立.

通过推广上一段的计算, 下面将证明, 如果 A 是分块上三角矩阵

$$A = \begin{bmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$

其中每个 A_j 都是 1×1 或 2×2 矩阵, 那么

10.27 $$\det A = (\det A_1)\cdots(\det A_m).$$

要证明这一点, 考虑 $\operatorname{perm} n$ 中的一个元素. 如果这个排列把对角线上对应于某个 1×1 块的指标移到任何其他位置, 那么这个排列在定义 $\det A$ 的和式 10.25 中对求和就没有贡献 (因为 A 是分块上三角矩阵). 对于对角线上的一个 2×2 块所对应的一对指标, 排列一定保持这对指标不变或者把它们互换; 否则, 这个排列在定义 $\det A$ 的和式 10.25 中对求和也没有贡献 (因为 A 是分块上三角矩阵). 根据这些观察和 2×2 矩阵的行列式公式 10.26 可得 10.27. 特别地, 如果 V 是实向量空间, $T\in\mathcal{L}(V)$, 并且取 V 的一个基使得关于此基 $\mathcal{M}(T)$ 是形如 9.10 的分块上三角矩阵, 而且对角上都是 1×1 或者 2×2 块, 那么 $\det T = \det\mathcal{M}(T)$.

推导行列式的性质足以写一整本书. 幸好我们只需要一些基本性质.

我们的目标是要证明, 对每个 $T \in \mathcal{L}(V)$ 和 V 的每个基都有 $\det T = \det \mathcal{M}(T)$. 为此, 需要给出行列式的一些性质. 下面的引理是其中的第一个性质.

10.28 引理: *假设 $\boldsymbol{A}$ 是一个方阵. 如果 $\boldsymbol{B}$ 是通过互换 $\boldsymbol{A}$ 的两列所得到的矩阵, 那么*

$$\det \boldsymbol{A} = -\det \boldsymbol{B}.$$

证明: 设 $\boldsymbol{A}$ 由 10.24 给出, 并且 $\boldsymbol{B}$ 是通过互换 $\boldsymbol{A}$ 的两列得到的. 考虑 $\det \boldsymbol{A}$ 的定义中的和式 10.25 以及 $\det \boldsymbol{B}$ 的定义中的相应和式. 这两个和式中出现的那些 a 的乘积是相同的, 只是相应的排列不同, 而且 $\det \boldsymbol{B}$ 的排列是通过互换 $\det \boldsymbol{A}$ 的相应排列中的两个元素得到的, 所以 $\det \boldsymbol{B}$ 的排列的符号等于$\det \boldsymbol{A}$ 的相应排列的符号乘以 -1 (参见 10.23). 因此 $\det \boldsymbol{A} = -\det \boldsymbol{B}$. ■

如果 $T \in \mathcal{L}(V)$, 并且 T (关于某个基) 的矩阵具有两个相同的列, 那么 T 不是单的, 于是 $\det T = 0$. 虽然对下一个引理的这个解释看似合理, 但却不能当成证明, 因为我们现在还不知道 $\det T = \det \mathcal{M}(T)$.

10.29 引理: *如果方阵 $\boldsymbol{A}$ 有两个列是相同的, 那么 $\det \boldsymbol{A} = 0$.*

证明: 设方阵 $\boldsymbol{A}$ 有两个列是相同的. 把 $\boldsymbol{A}$ 的这两个相同的列互换仍得到最初的矩阵 $\boldsymbol{A}$. 因此, 由 10.28 (取 $\boldsymbol{B} = \boldsymbol{A}$) 得 $\det \boldsymbol{A} = -\det \boldsymbol{A}$, 从而 $\det \boldsymbol{A} = 0$. ■

本节比较长, 所以我们暂停一段. 每章第一页出现的符号 ✻ 是用来占地方的装饰, 这是为了使每章的第一节从下一页开始. 第 1 章有 1 个这样的符号, 第 2 章有 2 个这样的符号, 依此类推. 这些符号随着章的增加而变小. 你可能还没有注意到, 各章开头的这些符号所占的面积之和是相同的. 例如, 第 10 章开头的每个符号的直径都等于第 1 章开头的符号的直径的 $1/\sqrt{10}$.

我们需要引入一种记号以便利用矩阵的列来表示矩阵. 如果 $\boldsymbol{A}$ 是 $n \times n$ 矩阵

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix},$$

那么我们可以把 $\boldsymbol{A}$ 的第 k 列看作一个 $n \times 1$ 矩阵

$$\boldsymbol{a}_k = \begin{bmatrix} a_{1,k} \\ \vdots \\ a_{n,k} \end{bmatrix}.$$

我们将把 $\boldsymbol{A}$ 写成

$$[\boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_n],$$

这里把 $\boldsymbol{a}_k$ 看作 $\boldsymbol{A}$ 的第 k 列. 采用这样的记号要注意, 具有两个下标的 $a_{j,k}$ 表示 $\boldsymbol{A}$ 的一个元素, 而具有一个下标的 $\boldsymbol{a}_k$ 表示 $\boldsymbol{A}$ 的一个列.

下一个引理说明, 把矩阵 $\boldsymbol{A}$ 的列重新排列, 行列式就变成了 $\boldsymbol{A}$ 的行列式乘以这个排列的符号.

有些教材把行列式定义为方阵的函数, 这个函数分别对每个列都是线性的, 并且满足 10.30 和 $\det I = 1$. 要证明这样的函数存在且唯一需要做大量的工作.

10.30 引理: 设 $\boldsymbol{A} = [\ \boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_n\ ]$ 是 $n \times n$ 矩阵. 如果 $(m_1, \cdots, m_n)$ 是一个排列, 那么

$$\det[\boldsymbol{a}_{m_1} \quad \cdots \quad \boldsymbol{a}_{m_n}] = (\operatorname{sign}(m_1, \cdots, m_n))\det \boldsymbol{A}.$$

证明: 假设 $(m_1, \cdots, m_n) \in \operatorname{perm} n$. 我们可以通过一系列步骤把矩阵 $[\ \boldsymbol{a}_{m_1} \quad \cdots \quad \boldsymbol{a}_{m_n}\ ]$ 变成 $\boldsymbol{A}$. 每一步我们都互换两列, 因此所得到的行列式等于前一个行列式乘以 -1 (参见 10.28). 所需要的步骤数等于把排列 $(m_1, \cdots, m_n)$ 变成排列 $(1, \cdots, n)$ 所需要的互换元素的次数, 为了完成证明, 只需注意到如果 $(m_1, \cdots, m_n)$ 的符号为 1, 那么这个次数是偶数; 如果 $(m_1, \cdots, m_n)$ 的符号是 -1, 那么这个次数是奇数 (由 10.23, 并注意到排列 $(1, \cdots, n)$ 的符号为 1). ■

令 $\boldsymbol{A} = [\ \boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_n\ ]$. 对于 $1 \leqslant k \leqslant n$, 把第 k 列之外的列都看作是固定的. 我们有

$$\det \boldsymbol{A} = \det[\boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_k \quad \cdots \quad \boldsymbol{a}_n],$$

并且我们可以把 $\det \boldsymbol{A}$ 看作第 k 列 $\boldsymbol{a}_k$ 的函数. 这个函数把 $\boldsymbol{a}_k$ 映成上面的行列式, 而且是从 $\mathbf{F}$ 上的 $n \times 1$ 矩阵所构成的向量空间到 $\mathbf{F}$ 的线性映射. 线性由 10.25 易得, 因为 10.25 中的每个和项都恰好包含 $\boldsymbol{A}$ 的第 k 列中的一个元素.

我们现在可以证明方阵的行列式的一个主要性质. 通过这个性质我们能够把算子的行列式和它的矩阵的行列式联系起来. 注意, 这个证明比关于迹的相应结果的证明要复杂得多 (参见 10.9).

10.31 定理: 如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是大小相同的方阵, 那么

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{BA}) = (\det \boldsymbol{A})(\det \boldsymbol{B}).$$

这个定理是由法国数学家比内 (Jacques Binet) 和柯西于 1812 年首先证明的.

证明: 设 $\boldsymbol{A} = [\ \boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_n\ ]$, 其中每个 $\boldsymbol{a}_k$ 都是 $\boldsymbol{A}$ 的一个 $n \times 1$ 的列. 令

$$\boldsymbol{B} = \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & & \vdots \\ b_{n,1} & \cdots & b_{n,n} \end{bmatrix} = [\boldsymbol{b}_1 \quad \cdots \quad \boldsymbol{b}_n],$$

其中每个 $\boldsymbol{b}_k$ 都是 $\boldsymbol{B}$ 的一个 $n \times 1$ 的列. 令 $\boldsymbol{e}_k$ 表示第 k 行的元素等于 1, 其余元素都等于 0 的 $n \times 1$ 矩阵. 注意到 $\boldsymbol{Ae}_k = \boldsymbol{a}_k$, $\boldsymbol{Be}_k = \boldsymbol{b}_k$. 进一步有, $\boldsymbol{b}_k = \sum_{m=1}^n b_{m,k}\boldsymbol{e}_m$.

首先证明 $\det(\boldsymbol{AB}) = (\det \boldsymbol{A})(\det \boldsymbol{B})$. 对矩阵乘法的定义稍加回忆即得 $\boldsymbol{AB} = [\ \boldsymbol{Ab}_1 \quad \cdots \quad \boldsymbol{Ab}_n\ ]$. 因此

$$\begin{aligned}
&\det(\boldsymbol{AB}) \\
&= \det[\ \boldsymbol{Ab}_1 \quad \cdots \quad \boldsymbol{Ab}_n\ ] \\
&= \det\left[\boldsymbol{A}\left(\sum\nolimits_{m_1=1}^n b_{m_1,1}\boldsymbol{e}_{m_1}\right) \cdots \boldsymbol{A}\left(\sum\nolimits_{m_n=1}^n b_{m_n,n}\boldsymbol{e}_{m_n}\right)\right] \\
&= \det\left[\sum\nolimits_{m_1=1}^n b_{m_1,1}\boldsymbol{Ae}_{m_1} \cdots \sum\nolimits_{m_n=1}^n b_{m_n,n}\boldsymbol{Ae}_{m_n}\right] \\
&= \sum\nolimits_{m_1=1}^n \cdots \sum\nolimits_{m_n=1}^n b_{m_1,1} \cdots b_{m_n,n} \det[\boldsymbol{Ae}_{m_1} \quad \cdots \quad \boldsymbol{Ae}_{m_n}],
\end{aligned}$$

其中最后一个等式是反复利用 det 作为第一列的函数是线性的这个性质得到的. 在上面的最后一个求和式中, 存在某个 $j \neq k$ 使得 $m_j = m_k$ 的所有项都可以忽略, 因为具有两个相同列的矩

阵的行列式等于 0 (由 10.29). 因此, 我们不需要对 $m_1, \cdots, m_n$ 的所有取值求和, 而只需对使得这些 m_j 具有不同值的排列求和, 其中每个 m_j 都取值于 $1, \cdots, n$. 也就是说

$$\begin{aligned}
&\det(\boldsymbol{AB}) \\
&= \sum_{(m_1,\cdots,m_n)\in \operatorname{perm} n} b_{m_1,1}\cdots b_{m_n,n} \det[\ \boldsymbol{A}\boldsymbol{e}_{m_1} \ \cdots \ \boldsymbol{A}\boldsymbol{e}_{m_n}\ ] \\
&= \sum_{(m_1,\cdots,m_n)\in \operatorname{perm} n} b_{m_1,1}\cdots b_{m_n,n} \big(\operatorname{sign}(m_1,\cdots,m_n)\big) \det \boldsymbol{A} \\
&= (\det \boldsymbol{A}) \sum_{(m_1,\cdots,m_n)\in \operatorname{perm} n} \big(\operatorname{sign}(m_1,\cdots,m_n)\big) b_{m_1,1}\cdots b_{m_n,n} \\
&= (\det \boldsymbol{A})(\det \boldsymbol{B}),
\end{aligned}$$

其中第二个等式由 10.30 得到.

上一段证明了 $\det(\boldsymbol{AB}) = (\det \boldsymbol{A})(\det \boldsymbol{B})$. 互换 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的角色得 $\det(\boldsymbol{BA}) = (\det \boldsymbol{B})(\det \boldsymbol{A})$, 即 $\det(\boldsymbol{BA}) = (\det \boldsymbol{A})(\det \boldsymbol{B})$. ■

现在我们可以证明, 算子的矩阵的行列式与计算这个矩阵所使用的基无关.

10.32 推论: *假设 $T \in \mathcal{L}(V)$. 如果 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ 和 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 都是 V 的基, 那么*

$$\det \mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)) = \det \mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)).$$

注意, 这个证明与关于迹的类似结果的证明相似 (参见 10.10).

证明: 假设 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ 和 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 都是 V 的基, 并设

$$\boldsymbol{A} = \mathcal{M}\big(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)\big),$$

那么

$$\begin{aligned}
\det \mathcal{M}(T, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)) &= \det(\boldsymbol{A}^{-1}(\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))\boldsymbol{A})) \\
&= \det((\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))\boldsymbol{A})\boldsymbol{A}^{-1}) \\
&= \det \mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)),
\end{aligned}$$

其中第一个等式由 10.3 得到, 第二个等式由 10.31 得到. 第三个等式完成了证明. ■

下面的定理是说, 算子的行列式等于该算子的矩阵的行列式. 这个定理并没有指明所用到的基, 这是因为根据上面的推论, 对于每个基来说, 算子的矩阵的行列式都相同.

10.33 定理: *若 $T \in \mathcal{L}(V)$, 则 $\det T = \det \mathcal{M}(T)$.*

证明: 令 $T \in \mathcal{L}(V)$. 上面已经注意到, 由 10.32 可知 $\det \mathcal{M}(T)$ 与 V 的基的选取无关. 因此, 要证明对 V 的每个基都有

$$\det T = \det \mathcal{M}(T),$$

只需证明上面的等式对 V 的某个基成立, 而这是已经证明过的 (在 230 页): 可取 V 的一个基使得关于此基 $\mathcal{M}(T)$ 是上三角矩阵 (如果 V 是复向量空间) 或者是适当的分块上三角矩阵 (如果 V 是实向量空间). ■

如果知道了复向量空间上一个算子的矩阵, 利用上面的定理, 不求算子的任何本征值就可以求出所有本征值的乘积. 例如, 考虑 $\mathbf{C}^5$ 上矩阵为

$$\begin{bmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

的算子. 没人知道这个算子的任何本征值的精确公式, 但是我们却知道它的本征值的乘积等于 -3, 这是因为上面矩阵的行列式等于 -3.

通过转换成矩阵行列式的语言, 利用上面的定理容易证明算子的行列式的一些有用的性质, 其中有些性质是已经证明过的, 或者是显然的. 我们在下一个推论中就来这么做.

10.34 推论: *若 $S, T \in \mathcal{L}(V)$, 则*

$$\det\,(ST) = \det\,(TS) = (\det\,S)(\det\,T).$$

证明: 假设 $S, T \in \mathcal{L}(V)$. 取 V 的任意一个基, 则有

$$\begin{aligned}\det(ST) &= \det \mathcal{M}(ST)\\ &= \det(\mathcal{M}(S)\mathcal{M}(T))\\ &= (\det \mathcal{M}(S))(\det\mathcal{M}(T))\\ &= (\det S)(\det T),\end{aligned}$$

其中第一个等式和最后一个等式由 10.33 得到, 第三个等式由 10.31 得到.

上一段证明了 $\det(ST) = (\det S)(\det T)$. 互换 S 和 T 的角色即得 $\det(TS) = (\det T)(\det S)$. 因为 **F** 中元素的乘法具有交换性, 所以上面的等式又可以写成 $\det(TS) = (\det S)(\det T)$. ■

§10.5 体 积

在这最后一章, 我们在引入行列式之前就已经证明了线性代数的基本结果. 虽然行列式作为更高等的课题的研究工具是有价值的, 但是它们在基础线性代数中并未发挥多少作用 (当该课题得到恰当处理时). 行列式在大学数学中确实有一个重要的应用, 即计算某些体积和积分. 在这最后一节中, 我们将利用所学习的线性代数知识来弄清楚行列式和这些应用之间的联系. 因此我们将利用线性代数来处理分析中的一部分内容.

大多数应用数学家认为, 行列式很难用于复杂的数值计算.

首先来看研究体积时需要用到的一些纯线性代数的结果. 回想一下, 内积空间中的等距同构是保持范数的算子. 下一个结果表明, 每个等距同构的行列式的绝对值都等于 1.

10.35 命题: *设 V 是内积空间. 若 $S \in \mathcal{L}(V)$ 是等距同构, 则 $|\det S| = 1$.*

证明: 设 $S \in \mathcal{L}(V)$ 是等距同构. 首先考虑 V 是复内积空间的情形. 此时 S 的所有本征值的绝对值都为 1 (由 7.37). 因此, S 的所有本征值 (计重数) 的乘积的绝对值也为 1. 也就是说, $|\det S| = 1$.

现在假设 V 是实内积空间, 那么 V 有一个规范正交基使得 S 关于此基有分块对角矩阵, 其中对角线上的每个块或者是由 1 或 -1 组成的 1×1 矩阵, 或者是如下形式的 2×2 矩阵

10.36
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

其中 $\theta\in(0,\pi)$(参见 7.38). 注意到每个形如 10.36 的矩阵的特征多项式的常数项都等于 1 (因为 $\cos^2\theta+\sin^2\theta=1$). 因此, S 的每个本征对的第二个坐标都等于 1. 于是 S 的行列式是一些 1 和一些 -1 的乘积. 特别地, $|\det S|=1$. ■

设 V 是实内积空间, $S\in\mathcal{L}(V)$ 是一个等距同构. 由上一个命题, S 的行列式等于 1 或者 -1. 注意到

$$\{\boldsymbol{v}\in V : S\boldsymbol{v}=-\boldsymbol{v}\}$$

是由 S 的相应于本征值 -1 的所有本征向量组成的 V 的子空间 (或者是子空间 $\{\mathbf{0}\}$, 如果 -1 不是 S 的本征值). 从几何的角度考虑, 我们可以说这个子空间是 S 的反向子空间. 仔细考察上一个命题的证明可知, 如果这个子空间的维数是偶数, 那么 $\det S=1$; 如果这个子空间的维数是奇数, 那么 $\det S=-1$.

实内积空间上的自伴算子没有本征对 (由 7.11). 因此, 实内积空间上的自伴算子的行列式等于它的所有本征值 (计重数) 的乘积 (当然, 这个结论对复向量空间上的任意算子都成立, 无论是不是自伴的).

回想一下, 如果 V 是内积空间, 并且 $T\in\mathcal{L}(V)$, 那么 T^*T 是正算子, 因此有唯一的正平方根, 记为 $\sqrt{T^*T}$ (参见 7.27 和 7.28). 因为 T^*T 是正的, 所以它的所有本征值都是非负的 (还是参见 7.27), 于是它的行列式是非负的. 因此在下面的讨论中不需要对 $\det T^*T$ 取绝对值.

10.37 推论: *假设 V 是内积空间. 若 $T\in\mathcal{L}(V)$, 则*

$$|\det T|=\det\sqrt{T^*T}.$$

本章的习题 24 提示了这个推论的另一个证明.

证明：假设 $T \in \mathcal{L}(V)$. 由极分解 (7.41), 存在等距同构 $S \in \mathcal{L}(V)$ 使得

$$T = S\sqrt{T^*T}.$$

因此

$$\begin{aligned}|\det T| &= |\det S|\det\sqrt{T^*T}\\ &= \det\sqrt{T^*T},\end{aligned}$$

其中第一个等式由 10.34 得到, 第二个等式由 10.35 得到. ■

设 V 是实内积空间, 并且 $T \in \mathcal{L}(V)$ 是可逆的. $\det T$ 是正的或者是负的. 仔细考察上面推论的证明可以帮助我们给各种情形都找出一个几何解释. 要看到这一点, 首先对正算子 T^*T 应用实谱定理 (7.13), 得到 V 的一个规范正交基 $(e_1, \cdots, e_n)$ 使得 $\sqrt{T^*T}e_j = \lambda_j e_j$, 其中 $\lambda_1, \cdots, \lambda_n$ 是 $\sqrt{T^*T}$ 的本征值, 按照重数重复. 因为每个 λ_j 都是正的, 所以 $\sqrt{T^*T}$ 根本不使任何向量反向. 现在考虑极分解

我们没有给出短语"反向"的正式定义, 因为这些解释只是为了帮助直观理解, 而不是严格的数学.

$$T = S\sqrt{T^*T},$$

其中 $S \in \mathcal{L}(V)$ 是一个等距同构, 则 $\det T = (\det S)(\det\sqrt{T^*T})$. 因此, $\det T$ 是正的还是负的取决于 $\det S$ 是正的还是负的. 前面我们已经看到, 这取决于 S 的反向子空间是偶数维的还是奇数维的. 因为 T 是 S 和一个根本不使任何向量反向的算子 (即 $\sqrt{T^*T}$) 的乘积, 所以我们有理由说 $\det T$ 是正的或者是负的取决于 T 使向量反向偶数次还是奇数次.

现在转向体积问题, 这里只考虑 (带有标准内积的) 实内积空间 $\mathbf{R}^n$. 我们给 $\mathbf{R}^n$ 的每个子集 Ω 指定一个 n 维体积, 记为 $\text{volume}\,\Omega$ (当 $n=2$ 时, 通常称之为面积而不是体积). 首先讨论立方体. 立方体的体积具有很好的直观. $\mathbf{R}^n$ 中边长为 r, 以 $(x_1, \cdots, x_n) \in \mathbf{R}^n$ 为顶点的**立方体 (cube)** 是集合

$$\{(y_1,\cdots,y_n)\in\mathbf{R}^n : x_j<y_j<x_j+r,\quad j=1,\cdots,n\};$$

你应该验证, 当 $n=2$ 时, 这是正方形, 当 $n=3$ 时, 这是我们熟悉的 3 维立方体. $\mathbf{R}^n$ 中边长为 r 的立方体的体积定义为 r^n. 定义任意集合 $\Omega\subset\mathbf{R}^n$ 的体积的思想就是把 Ω 写成很多小立方体的并集, 然后再把这些小立方体的体积相加. 用小立方体的并 (可能是无限并) 逼近 Ω 越精确, 对 Ω 的体积的估计就越好.

熟悉外测度的读者会看出这里的概念.

与其给体积下一个麻烦的严格定义, 倒不如使用体积的直观概念. 本书的目标是理解线性代数, 而体积的概念是属于分析的 (不过我们很快就会看到体积与行列式的紧密联系). 因此, 在本节的其余部分我们将借助体积的直观概念, 而不去做严格的推导. 但是在接下来所涉及的线性代数部分我们还是要保持一贯的严密. 这里所有关于体积的陈述都是正确的 —— 这里所给出的直观论证都可以利用分析手段转化成正式的严格证明.

对于 $T\in\mathcal{L}(V)$ 和 $\Omega\subset\mathbf{R}^n$, 定义 $T(\Omega)$为

$$T(\Omega)=\{T\boldsymbol{x}:\boldsymbol{x}\in\Omega\}.$$

我们的目标是利用 T 和 Ω 的体积来找出 $T(\Omega)$ 的体积公式. 先来看一个简单的例子. 设 $\lambda_1,\cdots,\lambda_n$ 是正数. 定义 $T\in\mathcal{L}(\mathbf{R}^n)$ 为 $T(x_1,\cdots,x_n)=(\lambda_1x_1,\cdots,\lambda_nx_n)$. 如果 Ω 是 $\mathbf{R}^n$ 中的一个边长为 r 的立方体, 那么 $T(\Omega)$ 是 $\mathbf{R}^n$ 中的一个边长为 $\lambda_1r,\cdots,\lambda_nr$ 的盒子. 这个盒子的体积为 $\lambda_1\cdots\lambda_nr^n$, 而 Ω 的体积为 r^n. 因此这个特定的 T 作用在一个立方体上所得到的体积是原来的 $\lambda_1\cdots\lambda_n$ 倍, 即 $\det T$ 倍.

同上, 假设 $\lambda_1,\cdots,\lambda_n$ 都是正数. 现在假设 $(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)$ 是 $\mathbf{R}^n$ 的规范正交基, T 是 $\mathbf{R}^n$ 上的算子, 使得 $T\boldsymbol{e}_j=\lambda_j\boldsymbol{e}_j$, $j=1,\cdots,n$. 在 $(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)$ 是 $\mathbf{R}^n$ 的标准基的特殊情况下, 这个算子与上一段所定义的算子相同. 甚至对任意的规范正交基 $(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)$, 这个算子与上一段所定义的算子都具有相同的性质 —— 它把第 j 个基向量乘以一个因子 λ_j. 因此, 我们有理由假设, 这个算子也把体积乘以了一个因子 $\lambda_1\cdots\lambda_n$, 即乘以了 $\det T$.

在介绍本节的主要结果之前, 我们还需要一些材料. 设 $S \in \mathcal{L}(\mathbf{R}^n)$ 是任意一个等距同构. 对于 $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{R}^n$, 我们有

$$\begin{aligned}\|S\boldsymbol{x} - S\boldsymbol{y}\| &= \|S(\boldsymbol{x} - \boldsymbol{y})\| \\ &= \|\boldsymbol{x} - \boldsymbol{y}\|.\end{aligned}$$

也就是说, S 不改变两点之间的距离. 你能想象得到, 这意味着 S 不改变体积. 具体地, 若 $\Omega \subset \mathbf{R}^n$, 则 $\operatorname{volume} S(\Omega) = \operatorname{volume} \Omega$.

现在我们可以给出伪证明: 算子 $T \in \mathcal{L}(\mathbf{R}^n)$ 的作用使体积变成了原来的 $|\det T|$ 倍.

10.38 定理: 若 $T \in \mathcal{L}(\mathbf{R}^n)$, 则

$$\operatorname{volume} T(\Omega) = |\det T|\,(\operatorname{volume} \Omega), \quad \Omega \subset \mathbf{R}^n.$$

证明: 首先, 考虑 $T \in \mathcal{L}(\mathbf{R}^n)$ 是正算子的情形. 设 $\lambda_1, \cdots, \lambda_n$ 是 T 的所有本征值, 按照重数重复. 每个本征值都是非负数 (参见 7.27). 由实谱定理 (7.13), V 有规范正交基 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$, 使得对每个 j 都有 $T\boldsymbol{e}_j = \lambda_j \boldsymbol{e}_j$. 如上面的讨论, 这说明 T 的作用使体积变成了原来的 $|\det T|$ 倍.

现在假设 $T \in \mathcal{L}(\mathbf{R}^n)$ 是任意一个算子. 由极分解 (7.41), 存在一个等距同构 $S \in \mathcal{L}(V)$ 使得

$$T = S\sqrt{T^*T}.$$

若 $\Omega \subset \mathbf{R}^n$, 则 $T(\Omega) = S\big(\sqrt{T^*T}(\Omega)\big)$. 因此

$$\begin{aligned}\operatorname{volume} T(\Omega) &= \operatorname{volume} S(\sqrt{T^*T}(\Omega)) \\ &= \operatorname{volume} \sqrt{T^*T}(\Omega) \\ &= (\det \sqrt{T^*T})(\operatorname{volume} \Omega) \\ &= |\det T|(\operatorname{volume} \Omega),\end{aligned}$$

其中第二个等式成立是因为等距同构 S 不改变体积 (如上面的讨论), 第三个等式成立是根据前面的讨论 (应用于正算子 $\sqrt{T^*T}$), 第四个等式成立是根据 10.37. ■

上面的定理导致了行列式在重积分的变量替换公式中的出现. 要说明这一点, 我们还是要使用含糊而直观的语言. 如果 $\Omega \subset \mathbf{R}^n$, 并且 f 是 Ω 上的一个实值函数 (不必是线性的), 那么 f 在 Ω 上的**积分 (integral)**, 记为 $\int_\Omega f$或者 $\int_\Omega f(x)\mathrm{d}x$, 定义如下: 将 Ω 分成足够小的小块使得 f 在每个小块上几乎都是常数; 在每个小块上用 f 的值 (几乎是常数) 乘以这个小块的体积, 然后再对所有的小块求和, 就得到了积分的一个近似. 对 Ω 的分块越细, 这个近似就越精确. Ω 实际上应该是一个适当的集合 (例如, 开集或可测集), 并且 f 也应该是一个适当的函数 (例如, 连续的或可测的), 但是我们不必担心这些技术问题. 注意到 $\int_\Omega f(x)\mathrm{d}x$ 中的 x 应该是一个哑变量, 可以用其他任何符号代替.

固定集合 $\Omega \subset \mathbf{R}^n$ 和函数 (不必是线性的) $\sigma : \Omega \to \mathbf{R}^n$. 我们将用 σ 对积分做变量替换. 在做变量替换之前需要定义 σ 的**导数 (derivative)**, 这个概念用到了线性代数. 对于 $x \in \Omega$, σ 在 x 点的导数是一个算子 $T \in \mathcal{L}(\mathbf{R}^n)$, 使得

$$\lim_{y \to 0} \frac{\|\sigma(x+y) - \sigma(x) - Ty\|}{\|y\|} = 0.$$

如果 $n = 1$, 那么这种意义下的导数就是乘以一元微积分中通常意义下的导数所诱导的 $\mathbf{R}$ 上的算子.

如果存在满足上式的算子 $T \in \mathcal{L}(\mathbf{R}^n)$, 则称 σ 在 x 点是**可微的 (differentiable)**. 如果 σ 在 x 点是可微的, 那么存在唯一的算子 $T \in \mathcal{L}(\mathbf{R}^n)$ 满足上面的等式 (我们将略去其证明). 这个算子 T 记为 $\sigma'(x)$. 直观地说, 我们的想法是, 对固定的 x 和很小的 $\|y\|$, $\sigma(x) + \big(\sigma'(x)\big)(y)$ 是 $\sigma(x+y)$ 的一个很好的近似 (注意到 $\sigma'(x) \in \mathcal{L}(\mathbf{R}^n)$, 所以这种写法是有意义的). 注意, 对固定的 x, 加上 $\sigma(x)$ 不改变体积. 所以, 如果 Γ 是 Ω 的一个包含 x 的小子集, 那么 $\operatorname{volume}\sigma(\Gamma)$ 近似于 $\operatorname{volume}\big(\sigma'(x)\big)(\Gamma)$.

因为 σ 是从 Ω 到 $\mathbf{R}^n$ 的函数, 所以可以写成

$$\sigma(x) = (\sigma_1(x), \cdots, \sigma_n(x)),$$

其中每个 σ_j 都是从 Ω 到 $\mathbf{R}$ 的函数. σ_j 对第 k 个坐标的偏导数记为 $D_k\sigma_j$. 求这个偏导数在点 $x \in \Omega$ 的值得 $D_k\sigma_j(x)$. 如果 σ 在 x 点是可微的, 那么 $\sigma'(x)$ 关于 $\mathbf{R}^n$ 的标准基的矩阵的第 j 行第 k 列元素为 $D_k\sigma_j(x)$ (我们将略去其证明). 也就是说,

$$\textbf{10.39} \quad \mathcal{M}(\sigma'(x)) = \begin{bmatrix} D_1\sigma_1(x) & \cdots & D_n\sigma_1(x) \\ \vdots & & \vdots \\ D_1\sigma_n(x) & \cdots & D_n\sigma_n(x) \end{bmatrix}.$$

设 σ 在 Ω 中的每一点都是可微的, 并且 σ 在 Ω 上是单的. 令 f 是定义在 $\sigma(\Omega)$ 上的实值函数. 令 $x \in \Omega$, 并且 Γ 是 Ω 的一个包含 x 的小子集. 上面注意到

$$\text{volume } \sigma(\Gamma) \approx \text{volume}(\sigma'(x))(\Gamma),$$

其中符号 $\approx$表示约等于. 由 10.38 即得

$$\text{volume } \sigma(\Gamma) \approx |\det \sigma'(x)|(\text{volume } \Gamma),$$

令 $y = \sigma(x)$. 在上式左端乘以 $f(y)$, 右端乘以 $f(\sigma(x))$ (因为 $y = \sigma(x)$, 所以这两个量是相等的), 得到

$$\textbf{10.40} \quad f(y)\text{volume } \sigma(\Gamma) \approx f(\sigma(x))|\det \sigma'(x)|(\text{volume } \Gamma).$$

现在, 把 Ω 分成许多小块, 并且把 10.40 对所有的小块求和, 得到

$$\textbf{10.41} \quad \int_{\sigma(\Omega)} f(y)\mathrm{d}y = \int_{\Omega} f(\sigma(x))|\det \sigma'(x)|\mathrm{d}x.$$

这个公式就是我们想要的结果, 称为变量替换公式, 因为你可以把 $y = \sigma(x)$ 看作一个变量替换.

进行变量替换的关键是要像 10.41 右端那样包含因子 $|\det \sigma'(x)|$. 最后, 通过两个重要的例子来说明这一点. 当 $n = 2$ 时, 可以利用由极坐标诱导的变量替换. 在这种情况下, σ 定义为

如果你不熟悉极坐标和球坐标可以跳过本节的其余部分.

$$\sigma(r, \theta) = (r\cos\theta, r\sin\theta),$$

熟悉极坐标的人显然都知道为什么我们在这里使用 r, θ 作为坐标, 而不用 x_1, x_2 (而对其他人来说却很神秘). 对于 σ 的这个选取, 你应该验证, 相应于 10.39 的偏导数的矩阵是

$$\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix},$$

上面矩阵的行列式等于 r, 这就解释了在利用极坐标计算积分时为什么需要因子 r.

最后, 当 $n=3$ 时, 可以利用由球坐标诱导的变量替换. 在这种情况下, σ 定义为

$$\sigma(\rho,\varphi,\theta)=(\rho\sin\varphi\cos\theta,\rho\sin\varphi\sin\theta,\rho\cos\varphi),$$

熟悉球坐标的人显然都知道为什么我们在这里使用 ρ,θ,φ 作为坐标, 而不用 x_1,x_2,x_3 (而对其他人来说却很神秘). 对于 σ 的这个选取, 你应该验证, 相应于 10.39 的偏导数的矩阵是

$$\begin{bmatrix} \sin\varphi\cos\theta & \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta \\ \sin\varphi\sin\theta & \rho\cos\varphi\sin\theta & \rho\sin\varphi\cos\theta \\ \cos\varphi & -\rho\sin\varphi & 0 \end{bmatrix},$$

你还应该验证, 上面矩阵的行列式等于 $\rho^2\sin\varphi$, 这就解释了在利用球坐标计算积分时为什么需要因子 $\rho^2\sin\varphi$.

习　　题

1. 设 $T \in \mathcal{L}(V)$, 并设 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 是 V 的基. 证明: $\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n))$ 是可逆的当且仅当 T 是可逆的.

2. 证明: 如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是大小相同的方阵, 并且 $\boldsymbol{AB} = \boldsymbol{I}$, 那么 $\boldsymbol{BA} = \boldsymbol{I}$.

3. 设 $T \in \mathcal{L}(V)$ 关于 V 的每个基的矩阵都相同. 证明 T 是恒等算子的标量倍.

4. 设 $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ 和 $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)$ 都是 V 的基, 并设算子 $T \in \mathcal{L}(V)$ 使得对 $k = 1, \cdots, n$ 都有 $T\boldsymbol{v}_k = \boldsymbol{u}_k$. 证明

$$\mathcal{M}(T, (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)) = \mathcal{M}(I, (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n)).$$

5. 证明: 如果 $\boldsymbol{B}$ 是一个复方阵, 那么存在可逆复方阵 $\boldsymbol{A}$ 使得 $\boldsymbol{A}^{-1}\boldsymbol{BA}$ 是上三角矩阵.

6. 举出实向量空间 V 及 $T \in \mathcal{L}(V)$ 的例子, 使得 $\operatorname{trace}(T^2) < 0$.

7. 设 V 是实向量空间, $T \in \mathcal{L}(V)$, 并且 V 有一个由 T 的本征向量组成的基. 证明 $\operatorname{trace}(T^2) \geqslant 0$.

8. 设 V 是内积空间, 并且 $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{L}(V)$. 定义 $T \in \mathcal{L}(V)$ 为 $T\boldsymbol{u} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle \boldsymbol{w}$. 求 $\operatorname{trace} T$.

9. 证明: 如果 $P \in \mathcal{L}(V)$ 满足 $P^2 = P$, 那么 $\operatorname{trace} P$ 是非负整数.

10. 证明: 如果 V 是内积空间, 并且 $T \in \mathcal{L}(V)$, 那么

$$\operatorname{trace} T^* = \overline{\operatorname{trace} T}.$$

11. 设 V 是内积空间. 证明: 如果 $T \in \mathcal{L}(V)$ 是正算子, 并且 $\operatorname{trace} T = 0$, 那么 $T = 0$.

12. 设算子 $T \in \mathcal{L}(\mathbf{C}^3)$ 的矩阵为

$$\begin{bmatrix} 51 & -12 & -21 \\ 60 & -40 & -28 \\ 57 & -68 & 1 \end{bmatrix}.$$

并且已知 -48 和 24 是 T 的本征值. 不用计算机也不用笔, 求 T 的第三个本征值.

13. 证明或举反例: 如果 $T \in \mathcal{L}(V)$, 并且 $c \in \mathbf{F}$, 那么

$$\operatorname{trace}(cT) = c \operatorname{trace} T.$$

14. 证明或举反例: 如果 $S, T \in \mathcal{L}(V)$, 那么

$$\operatorname{trace}(ST) = (\operatorname{trace} S)(\operatorname{trace} T).$$

15. 设 $T \in \mathcal{L}(V)$. 证明: 如果对所有 $S \in \mathcal{L}(V)$ 都有 $\operatorname{trace}(ST) = 0$, 那么 $T = 0$.

16. 设 V 是内积空间, 并且 $T \in \mathcal{L}(V)$. 证明: 如果 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ 是 V 的规范正交基, 那么

$$\operatorname{trace}(T^*T) = \|T\boldsymbol{e}_1\|^2 + \cdots + \|T\boldsymbol{e}_n\|^2.$$

证明上式右端与 V 的规范正交基 $(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$ 的选取无关.

17. 设 V 是复内积空间, $T \in \mathcal{L}(V)$, 并设 $\lambda_1, \cdots, \lambda_n$ 是 T 的本征值, 按照重数重复. 设

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix}$$

是 T 关于 V 的某个规范正交基的矩阵. 证明

$$|\lambda_1|^2 + \cdots + |\lambda_n|^2 \leqslant \sum_{k=1}^{n} \sum_{j=1}^{n} |a_{j,k}|^2.$$

18. 设 V 是内积空间. 证明

$$\langle S, T \rangle = \operatorname{trace}(ST^*)$$

定义了 $\mathcal{L}(V)$ 上的一个内积.

习题 19 对无限维内积空间不成立, 从而引出所谓的亚正规算子, 亚正规算子的理论发展已经比较成熟.

19. 设 V 是内积空间, 并且 $T \in \mathcal{L}(V)$. 证明: 如果对每个 $\boldsymbol{v} \in V$ 都有

$$\|T^*\boldsymbol{v}\| \leqslant \|T\boldsymbol{v}\|,$$

那么 T 是规范的.

20. 证明或举反例: 如果 $T \in \mathcal{L}(V)$, 并且 $c \in \mathbf{F}$, 那么

$$\det(cT) = c^{\dim V} \det T.$$

21. 证明或举反例: 如果 $S, T \in \mathcal{L}(V)$, 那么

$$\det(S+T) = \det S + \det T.$$

22. 设 $\boldsymbol{A}$ 是分块上三角矩阵

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & & * \\ & \ddots & \\ 0 & & \boldsymbol{A}_m \end{bmatrix},$$

其中对角线上的每个 $\boldsymbol{A}_j$ 都是方阵. 证明

$$\det \boldsymbol{A} = (\det \boldsymbol{A}_1) \cdots (\det \boldsymbol{A}_m).$$

23. 设 $\boldsymbol{A}$ 是一个 $n \times n$ 的实矩阵. 令 $S \in \mathcal{L}(\mathbf{C}^n)$ 是 $\mathbf{C}^n$ 上的算子, 矩阵为 $\boldsymbol{A}$, 而 $T \in \mathcal{L}(\mathbf{R}^n)$ 是 $\mathbf{R}^n$ 上的算子, 矩阵也为 $\boldsymbol{A}$. 证明 $\operatorname{trace} S = \operatorname{trace} T$, 并且 $\det S = \det T$.
24. 设 V 是内积空间, 并且 $T \in \mathcal{L}(V)$. 证明

$$\det T^* = \overline{\det T}.$$

由此证明 $|\det T| = \det\sqrt{T^*T}$, 这给出了一个与 10.37 不同的证明.
25. 令 a, b, c 为正数. 求一个集合 $\Omega \subset \mathbf{R}^3$ 和一个算子 $T \in \mathcal{L}(\mathbf{R}^3)$, 使得 $T(\Omega)$ 等于椭球

$$\left\{(x,y,z) \in \mathbf{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} < 1\right\}.$$

假设已知 Ω 的体积, 求上面椭球的体积.

符号索引

索　引

H

J

K

L

M

N

O

P

Q

S

版 权 声 明